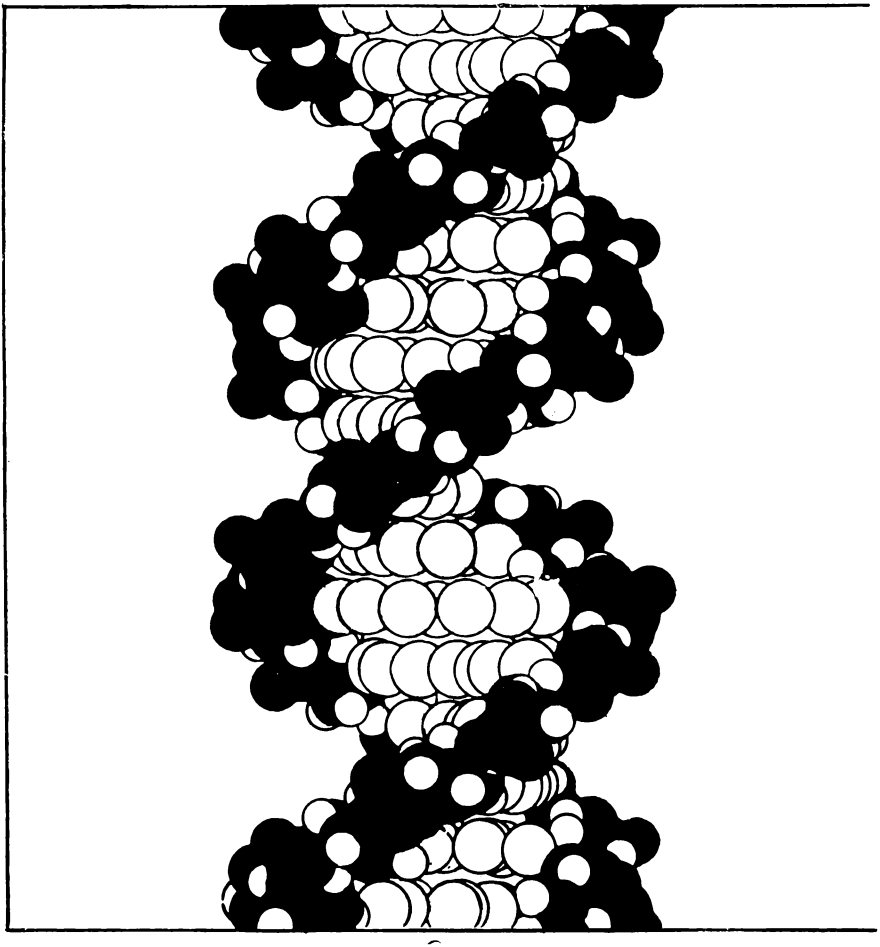
biophysics

M. V. Volkenshtein

- ☐ Tasks and Methods of Biophysics
- ☐ Chemical Foundations of Biophysics
- ☐ Physics of Macromolecules
- ☐ The Physics of Proteins
- ☐ Methods of Investigation of the Structure of Biopolymers
- ☐ Physics of Enzymes
- ☐ Physics of Nucleic Acids
- ☐ Physics of Protein Biosynthesis
- ☐ Nonequilibrium Thermodynamics in Biology
- ☐ Physics of Membranes
- ☐ Physics of the Nervous Impulse
- ☐ Mechanochemical Processes
- ☐ Bioenergetics of the Respiratory Chain
- ☐ Photobiological Processes
- ☐ Modelling of Dynamic Biological Processes
- ☐ Periodic Chemical and Biological Processes
- ☐ Problems of Biological Development

MIR Publishers
Moscow

ИЗДАТЕЛЬСТВО
МИР



biophysics

M. V. Volkenshtein

Translated from the Russian
by Artavaz Beknazarov

Mir Publishers
Moscow

БИОФИЗИКА
М. В. Волькенштейн
Издательство «Наука» Москва

First published 1983
Revised from the 1981 Russian edition

На английском языке

© Издательство «Наука», 1981
© English translation, Mir Publishers, 1983

Preface to the English Edition

This book is intended for students and post-graduates—biologists and physicists—who study biophysics. The book has emerged from the lectures given by the author for many years at the Moscow Physico-Technical Institute and from two monographs published earlier. The first of these monographs, *Molecular Biophysics*, was published in Russian in 1975 and in English by Academic Press in 1977. The second, *General Biophysics*, was issued in Russian in 1978 and is being prepared for publication by Academic Press. A considerable part of the book has been written anew. At the same time, the above two monographs are recommended to the reader for a more thorough study of the subject. They are provided with extensive bibliography.

In my opinion, biophysics is not a subsidiary branch of biology but a legitimate part of the physical sciences. Biophysics is the physics of living things. Accordingly, an attempt has been made to divorce biophysics from physiology, though these fields overlap to a considerable extent. Though the book is devoted mainly to the theoretical problems of biophysics, it also deals with the most important experimental findings.

I am well aware of the difficulties associated with the writing of an encyclopedic course of biophysics by one author. Today, biophysics is a very vast, complicated and diversified branch of natural science and one cannot be, so to say, a jack-of-all-trades, a specialist equally well versed in all the problems dealt with in biophysics.

This unavoidably leads to deficiencies of exposition. The author will appreciate all critical remarks from the reader.

The author expresses his gratitude to Professors L. A. Blumenfeld and A. B. Rubin who read the manuscript and made valuable suggestions.

M. Volkenshtein

Moscow, March 1982

Contents

Preface 5

Chapter 1. Goals and Methods of Biophysics 11
1.1. The Place of Biophysics in Natural Sciences 11
1.2. Physics and Biology 13
1.3. Living and Inanimate Nature 15
1.4. Biological Individuality 17
1.5. Finalism and Causality 18
1.6. The Properties of Open Systems 19
1.7. Scope and Methods of Biophysics 22

Chapter 2. Chemical Foundations of Biophysics 25
2.1. Chemistry and Biology 25
2.2. Amino Acids 27
2.3. Electrolytes 29
2.4. The Composition and Primary Structure of Proteins 34
2.5. Nucleic Acids 36
2.6. Adenylates 43
2.7. The Chirality of Biological Molecules 45
2.8. Carbohydrates and Lipids 49
2.9. Cofactors, Vitamins, and Hormones 51
2.10. The Main Biochemical Processes 55
2.11. Strong and Weak Interactions 58

Chapter 3. Physics of Macromolecules 63
3.1. Macromolecules and Rubber Elasticity 63
3.2. Internal Rotation and Rotational Isomerism 66
3.3. The Rotational-Isomeric Theory of Macromolecules 71
3.4. The Macromolecule as a Cooperative System 78
3.5. The Coil and the Globule 82
3.6. Methods of Investigation of Macromolecules 85
3.7. Polyelectrolytes 91

Chapter 4. Physics of Proteins 95
4.1. The Goals of Protein Physics 95
4.2. Conformations of the Polypeptide Chain 97
4.3. The Hydrogen Bond and the Structure of Water 103
4.4. Helix-Coil Transitions 108
4.5. The Protein Globule and Hydrophobic Interactions 114
4.6. Relationship Between the Primary and the Three-Dimensional Structure of Proteins 119
4.7. The Structure and Stability of the Globule 125
4.8. Antibodies and Antigens 132
4.9. Fibrous Proteins 138

Chapter 5. Methods of Investigation of the Structure of Biopolymers . . 142
5.1. X-Ray Diffraction Analysis 142
5.2. Diffuse Scattering of X-Rays by Solutions of Biopolymers . . 148
5.3. Methods of Nuclear Physics 150
5.4. Electronic Absorption Spectra 153
5.5. Luminescence 157
5.6. Optical Activity 161
5.7. Optical Activity of Biopolymers 169
5.8. Magnetic Optical Activity 174
5.9. Vibrational Spectra 178
5.10. Nuclear Magnetic Resonance and Electron Paramagnetic Resonance 182

Chapter 6. Physics of Enzymes 188
6.1. Chemical Kinetics and Catalysis 188
6.2. Kinetics of Simple Enzymatic Reactions 193
6.3. Chemical Aspects of Enzyme Action 198
6.4. Conformational Properties of Enzymes 205
6.5. Physics of Enzyme-Substrate Interactions 209
6.6. Electronic-Conformational Interactions 212
6.7. Cooperative Properties of Enzymes 217
6.8. Myoglobin and Hemoglobin 224

Chapter 7. Physics of Nucleic Acids 236
7.1. Molecular Biology and Physics 236
7.2. The Structure of Nucleic Acids 239
7.3. Intramolecular Interactions in the Double Helix 247
7.4. Thermodynamics of Melting of the Double Helix 250
7.5. Kinetics of Unwinding of the Double Helix 260
7.6. Interaction of the Double Helix with Small Molecules and Ions 265
7.7. Replication of DNA 268

Chapter 8. Physics of Protein Biosynthesis 275
8.1. The Problem of the Genetic Code 275
8.2. Protein Biosynthesis 280
8.3. Transcription and Reverse Transcription 284

8.4. Transfer RNA 286
8.5. Translation 289
8.6. Deciphering of the Genetic Code and Its Meaning 294
8.7. Mutations 301
8.8. Regulation of Genes 306

Chapter 9. Nonequilibrium Thermodynamics in Biology 313
9.1. Information and Entropy 313
9.2. Nonequilibrium Processes 319
9.3. Coupling of Fluxes 322
9.4. Coupling of Chemical Reactions 325
9.5. Steady States of Linear Systems 329
9.6. Coupling of Chemical Reactions to Transport of Matter 335
9.7. Far-from-Equilibrium Processes 339

Chapter 10. Physics of Membranes 346
10.1. Cell Membranes 346
10.2. The Structure of Membranes 348
10.3. Conformational Properties of Membranes 351
10.4. Passive Membrane Transport 355
10.5. Active Membrane Transport 361
10.6. Transport of Charged Particles Through Membranes 368
10.7. Molecular Reception 372

Chapter 11. Physics of the Nerve Impulse 377
11.1. The Axon and the Nerve Impulse 377
11.2. Propagation of the Nerve Impulse 388
11.3. Generation of the Nerve Impulse 394
11.4. Ionic Channels 399
11.5. Synaptic Transmission 402

Chapter 12. Mechanochemical Processes 406
12.1. Thermodynamics of Mechanochemical Processes 406
12.2. The Structure of the Muscle and Muscle Proteins 411
12.3. The Chemistry and Physics of the Muscle 417
12.4. The Theory of Muscular Contraction 423
12.5. The Kinetic Properties of the Muscle 431
12.6. Mechanochemical Systems 433
12.7. Biomechanics 437

Chapter 13. Bioenergetics of the Respiratory Chain 440
13.1. Biological Oxidation 440
13.2. The Structure and Properties of Mitochondria 446
13.3. Chemiosmotic Coupling 449
13.4. Electronic-Conformational Interactions 457
13.5. Cytochrome *c* 463

Chapter 14. Photobiological Processes 466
14.1. Photosynthesis 466
14.2. Two Photochemical Systems 472
14.3. Chloroplasts 477
14.4. The Mechanism of Photosynthesis 481
14.5. Vision 485
14.6. The Molecular Mechanism of Photoreception 489
14.7. Bacteriorhodopsin 497

Chapter 15. Modelling of Dynamic Biological Processes 501
15.1. Dynamic Order 501
15.2. Physico-Mathematical Foundations of the Dynamics of Nonlinear Processes 504
15.3. Lotka-Volterra Models 515
15.4. Autocatalytic Systems 521
15.5. Phase Transitions 527
15.6. Stochastic Processes 532
15.7. Dynamics and Regulation 535

Chapter 16. Periodic Chemical and Biological Processes 538
16.1. Introduction 538
16.2. Belousov-Zhabotinsky Reactions 539
16.3. Auto-oscillations in Glycolysis 547
16.4. Auto-oscillations in Photosynthesis 551
16.5. Nonlinear Dynamics of Membranes 554
16.6. Autowave Processes in the Cardiac Muscle 559

Chapter 17. Problems of Biological Evolution and Development 564
17.1 The Origin of Life 564
17.2. Modelling of Prebiotic Evolution 568
17.3. Game Models and Informational Aspects of Self-Organization 572
17.4. Hypercycles 575
17.5. Other Models of Prebiological Evolution 579
17.6. Biological Evolution 585
17.7. Ontogenesis 589
17.8. Immunity 596
17.9. Biological Evolution and Information Theory 602
17.10. Complexity and Evolution 611
Recommended Literature 619
Name Index 624
Subject Index 627

Chapter 1 Goals and Methods of Biophysics

1.1. The Place of Biophysics in Natural Sciences

We shall adopt the definition of physics as a science dealing with the study of the structure and properties of concrete forms of matter—substances and fields, and of the forms of existence of matter—space and time. This definition does not imply any discrimination between living organisms and inanimate matter. It does not reduce the entire natural science to physics, but it indicates that the basic theoretical foundations of any branch of natural science are of a physical nature. Such foundations have already been discovered in chemistry and today we know that chemistry is concerned with the study of the structure of the electronic shells of atoms and molecules and of its change in the course of their interactions. Accordingly, theoretical chemistry is based at present on quantum and statistical mechanics, on thermodynamics and physical kinetics.

Biology is a science dealing with living organisms which are immeasurably more complicated than nonliving matter. Therefore, we have still to travel a long path before we can get a more penetrating insight into the deep-seated physical laws and principles governing biological phenomena.

Based on what has been said above, we define biophysics as the physics of life phenomena studied at all levels, from molecules and cells to the biosphere as a whole. Such a definition of biophysics contrasts with its being regarded as a subsidiary branch of biology and physiology. The content of biophysics is not necessarily associated with the application of physical instruments in a biological experiment. A clinical thermometer, an electrocardiograph, and a microscope are all physical instruments, but physicians and biologists who employ these devices are not at all concerned with biophysics.

A biophysical investigation commences with the physical formulation of a problem pertaining to living organisms. This means that such a problem is formulated on the basis of the general laws of physics and the atomic and molecular structure of substances.

Thereby, the ultimate goal of biophysics consists in substantiating theoretical biology. At the same time, biophysics is concerned with

the solution of numerous theoretical and practical (applied) problems of a more specific nature.

The formulation of the biophysical task is possible so far only in a limited number of cases. Living systems are so complex that biological knowledge is mostly insufficient for physical approaches to be worked out. Biology, however, is progressing at a very rapid rate and the development of biophysics is inseparably tied up with the advancement of biology.

Biophysics is a science of the 20th century. This is not to say that biophysical problems had not been solved at all before. Maxwell enunciated his theory of colour vision, Helmholtz measured the velocity of propagation of the nervous impulse. Many other examples could be cited. However, only in our time has biophysics switched over from the study of the physical properties of living organisms and of their physical stimulation (light, sound, electricity) to the investigation of the fundamental problems—heredity and variability, ontogenesis and phylogenesis, metabolism and bioenergetics. This has been made possible only as a result of the vigorous development of biology and biochemistry.

The objectives of biophysics are analogous to those of biology. They consist in understanding and interpreting life phenomena. Being a part of physics, biophysics is inseparable from biology. The biophysicist must have both physical and biological knowledge. For the work in the field of biophysics to be successful, a general understanding of living matter is desirable, which can be achieved through the knowledge of the fundamentals of zoology and botany, physiology and ecology. Physicists often neglect the descriptive sections of biology. The knowledge of zoology and botany is fundamentally significant; without Linnaeus the Darwin theory of evolution could not have originated.

Despite serious difficulties, modern biophysics has made great advances in the explanation of a number of biological phenomena. We have learnt a lot about the structure and activity of biologically functional molecules, about the properties and mechanisms of the function of cellular structures, such as membranes, bioenergetic organelles, mechanochemical systems. Physico-mathematical models of biological processes are being successfully constructed, including those of ontogenesis and phylogenesis. General theoretical approaches to life phenomena have been worked out, which are based on thermodynamics, information theory, the theory of automatic control. All these topics will be examined in this book in varying degrees of detail. In accordance with the definition of biophysics as the physics of life phenomena, we shall resort to physical laws rather than to the physiological classification. For example, the reception of external stimuli by sense organs is considered in various sections of the book—vision in a chapter devoted to photobiological phenom-

ena; the senses of hearing and touch in a chapter dealing with mechanochemical processes; the sense of smell is treated in connection with the physics of molecular recognition.

1.2. Physics and Biology

We may ask: Is contemporary physics sufficient for the solution of biological problems, for the substantiation of theoretical biology? Perhaps, biophysics will need a new, as yet non-existent physics? There occurred situations in the history of science, when the earlier developed theory reached its limits of applicability and it proved necessary to build up a fundamentally new set of concepts. It is exactly in this way that the theory of relativity and quantum mechanics emerged. The possibility is not excluded *a priori* that true biophysics must be built up on the basis of the yet unknown future physics, essentially different from modern physics.

We put aside the concept of *vitalism*, according to which biological phenomena are basically incomprehensible on the basis of physics and chemistry since there exists a certain mysterious "vital force", or entelechy, or a biological field, which are not amenable to a physical interpretation. The vital-force theory has nothing to do with science; it denies the unity of nature and, in the final run, leads to theology. Vitalistic conceptions are no longer existent today in modern science.

While discussing the possibilities of a physical interpretation of life phenomena, i.e., the impact of physics on the present-day and further development of biology, we should not forget about the influence of biology on physics. The law of conservation of energy, which is the first law of thermodynamics, was discovered by Mayer, Joule, and Helmholtz. As is known, Mayer's conceptions were based on observations of living organisms, human beings. It is less widely known that Helmholtz too used biological phenomena, being guided by a distinct antivitalistic conception. He wrote, "According to Stahl, the forces that are operating in a living body are in essence the physical and chemical forces of the organs and substances, but some living soul or life force inherent in the body is capable of binding or releasing their activity ... I have found that the Stahl theory ascribes to any living body the properties of a so-called *perpetuum mobile* (a perpetual-motion machine) ... Thus, I have run across the question of what relationships exist between the various natural forces if it is assumed that the *perpetuum mobile* is absolutely impossible...". Not only biophysics but physics too as a whole had to overcome the vital-force theory in the course of their development.

Let us consider the concepts of the interrelationship between physics and biology that have been worked out by the physicists of the twentieth century—Bohr and Schrödinger.

Bohr treated this problem on the basis of the *principle of complementarity*, a special case of which is the Heisenberg uncertainty principle. Bohr maintained that the biological laws proper are complementary to the laws obeyed by inanimate things. The physicochemical properties of a living organism and the life phenomena cannot be studied simultaneously—the cognition of one excludes that of the other. Life should be looked upon as the basic postulate of biology that does not lend itself to further analysis, just as the existence of a quantum of action forms the elementary basis of nuclear physics. Thus, Bohr believed that biological and physico-chemical investigations are complementary to each other, i.e., are incompatible, though not contradictory. This concept is not vitalistic since it denies the existence of a limit to the application of physics and chemistry for the solution of biological problems. Bohr wrote that no single result of a biological investigation can be described unambiguously unless the physical and chemical concepts are resorted to.

Bohr's view changed with the progress of modern biology. Later, he spoke of a complementarity between the physico-chemical considerations practically employed in biology and the concepts that are directly associated with the unity of the organism and go beyond the framework of physics and chemistry. The utilization of the concept of complementarity in biology was advocated by Bohr through reference not to the postulative nature of the life concept but to the extreme complexity of the organism as a whole system. In other words, not a fundamental but rather a practical, i.e., surmountable, complementarity is meant. This cannot be disputed.

In 1945 Schrödinger published a book called *What is Life?* which exerted a substantial impact on the development of biophysics and molecular biology. Several problems of fundamental significance are treated in detail in this book. The first of these is the thermodynamic foundations of life. At first glance there is a distinct contradiction between the evolution of an isolated physical system towards the state of maximum entropy, i.e., disorder (the second law of thermodynamics), and biological evolution which goes from the simple to the complex. Schrödinger stated that the living organism "feeds on negative entropy". This means that the organisms and the biosphere as a whole are not isolated but open systems which exchange both matter and energy with the outside world. The nonequilibrium state of an open system is maintained through the outflow of the entropy into the surroundings. The second problem is the general structural features of organisms. In Schrödinger's words, an organism is an *aperiodic crystal*, i.e., a highly ordered system, like a solid body, but devoid of the periodicity in the arrangement of cells, molecules, and atoms. This statement is valid for the structure of organisms, cells, and biological macromolecules (proteins, nucleic acids). As we

shall see, the concept of the aperiodicity of a crystal is essential to the treatment of life phenomena on the basis of information theory. The third problem is the conformity of biological phenomena to the laws of quantum mechanics. While discussing the results of the radiobiological investigations carried out by Timofeev-Ressovsky, Delbrück, and others, Schrödinger pointed out the quantum nature of radiation mutagenesis. At the same time, the applications of quantum mechanics in biology are not trivial since living organisms are essentially macroscopic. Schrödinger posed the question: "Why are atoms small?" It is obvious that this question is meaningless if no point of reference is specified. They are small in comparison with our measures of length—the metre, the centimetre. But these measures are determined by the dimensions of the human body. Hence, Schrödinger says, the question should be reformulated: why are atoms much smaller than the organisms? In other words, why are the organisms built up of a large numbers of atoms? Indeed, the number of atoms in the smallest bacterial cell *Mycoplasma laidlawii* is of the order of 10^9. The answer to the question is that the orderliness necessary for life is possible only in a macroscopic system; otherwise, the order would be destroyed by fluctuations. And, finally, Schrödinger posed the question of the stability of the substance of genes, which is made up of the light atoms, C, H, N, O, and P, during the lifetime of numerous generations. The answer to this question was later provided by molecular biology which established the double-helical structure of deoxyribonucleic acid (DNA).

We shall not dwell on the works of some physicists who maintained that biological phenomena do not conform to the laws of physics (Elsasser, Wigner, and others). The erroneousness of these works was disclosed in scientific literature.

There is every ground today to assert that modern physics is not approaching its limits of applicability for the treatment of biological phenomena. It is hardly probable that such limits will be reached in the future. On the contrary, the advancement of biophysics as a part of modern physics points to its unlimited possibilities. One has, of course, to introduce new physical conceptions but not new principles and laws.

1.3. Living and Inanimate Nature

Let us first define a living organism as an open, self-regulating, self-reproducing and developing heterogeneous system whose most important functional substances are biopolymers—proteins and nucleic acids. A living organism is a historical system in the sense that it is the result of the phylogenetic evolution and itself covers the path of ontogenetic development—from zygote to old age and death.

The conventional physics of inanimate matter has nothing to do with history. The electron, atom, and molecule are characterized by constant physical properties, irrespective of their origin. Of course, conventional physics deals with kinetic, dynamic processes. The individual past history of a physical body is not considered, however.

What has been said above does not mean that there are no historical problems in the physics of inanimate nature. The origin of living matter itself, its evolution, and the individual development of each species is a part of the development of the Universe as a whole, a part of the development of the solar system, a part of the development of the Earth. Hence, there is every reason to consider the similarity and differences between biophysics, on the one hand, and cosmology, astrophysics, and geophysics, on the other. This is instructive because these different branches of physics may enrich each other with unified approaches to the solution of historical problems.

According to modern concepts, the history of the Universe begins with a small cluster of high-density plasma. About 2×10^{10} years ago this plasma cluster began to expand explosively, electrons and nuclons emerging from photons and neutrinos; as the Universe cooled, light and then heavy atoms were formed. Evidently, the expansion of Universe is a continuous and irreversible process. The forces of gravity made possible the formation of stars and galaxies. On high gravitational compression the temperature of a star rises up to the moment when there occur thermonuclear processes. These processes are responsible for the evolution of stars, for catastrophic events (so-called cataclysms of nature), such as the appearance of supernovae. The Sun is a star which is in a certain stage of evolution; the formation of the planets is one of the consequences of the development of the sun. According to present-day findings, the Earth exists for about 4.5×10^9 years and the life on the Earth for about $3.5 \times \times 10^9$ years.

Cosmology and astrophysics are based on the theory of relativity and quantum (nuclear) physics, on thermodynamics and statistical mechanics. The evolution of stars is not linked up with their "struggle for existence", i.e., with their interactions with the environment and with competing interrelationships. This is a non-Darwinian evolution. Accordingly, it might seem that there is nothing common between biophysics and astrophysics.

There are, however, two points of similarity. First, the evolution of both the Universe and the biosphere involves an irreversible and continuous *generation of new information.* New information appears as a result of the memorizing of an accidental choice. Second, the irreversible development in both cases does mean the presence of *instabilities.* A new stage in the evolution of a star, a planet, the biosphere, biogeocenosis, a biological species or population arises as

a result of the instability of the preceding stage. This explains why the process of evolution is irreversible.

Thus, theoretical approaches based on information theory and the consideration of the stability of dynamical systems are basically common to the physics of living and inanimate things.

1.4. Biological Individuality

An essential feature of living organisms is their unlimited variety At present about 3×10^6 species of various living creatures are known to exist. The number of different individuals of multicellular plants or invertebrates cannot be estimated—it is exceedingly great. We cannot so far discern the individualities of the representatives of a particular strain of unicellular organisms, but, one may suppose, such specific features do exist. No two absolutely identical organisms exist on the earth. This is accounted for by the genetic variability, which is realized within very wide limits, and by the differences in interactions with the surroundings. The Darwinian evolution is tied up with variability, with the unlimited individualization of living organisms. Accordingly, scientific biology could not have existed without its descriptive branches—zoology and botany.

The situation is different with inanimate matter. The number of various atoms, including isotopes, is several hundreds. Today, about 10^2 elementary particles are known. The number of various atoms and simple molecules observed in the outer space is also of the order of 10^2 and one has not to deal with a larger number of elementary entities in investigations of the evolution of stars. The number of various compounds known to modern chemistry is of the order of 10^6, but in this case there are no limits at all. It is essential, however, that atoms and molecules of a particular substance are indistinguishable from one another.

Of course, in inanimate nature too there exists some kind of individualization at the macroscopic level. Real crystals of a given mineral differ from one another, since they have been grown under different conditions. There are no two absolutely identical stones rounded off by the sea tide. This variety, however, is of no prime importance to the development of inanimate things, which, though variable, display no heredity.

What chemical, molecular foundations could there be for the unlimited variety of living systems? They are determined by the macromolecular structure of genes and organisms. Polymeric chains, macromolecules do not obey the basic law of chemistry—the law of constant composition. Let us imagine a copolymer built up of two species of monomeric units. If the number of units in a macromolecular chain is equal to 100 (this is far from being a long chain), then the number of various chains containing two types of monomeric

units will be equal to $2^{100} \approx 10^{30}$. Thus, the macroscopic sample of such a copolymer may not contain two identical macromolecules.

As a matter of fact, biological variability is determined in the long run by the variety of genes, i.e., sufficiently long portions of DNA macromolecules, which are copolymers made up of four types of repeating units.

Like physics in general, biophysics is a quantitative science which makes wide use of the mathematical apparatus. Biology as such, primarily population genetics, has also to make recourse to mathematics. Serious difficulties are, however, encountered in a mathematical description of the individual features of organisms. Such a description may be only numerical but not analytical.

1.5. Finalism and Causality

The biological evolution is determined by the predominant survival of populations that are best adapted to the environmental conditions. Accordingly, the structure of a living organism is characterized by fitness and adaptation to a certain ecological niche. Therefore, in biology there naturally appears a finalistic interpretation of the phenomena studied. The growth of a zygote into an adult organism can be described, using the concept of design or purpose: the purpose of the development is the creation of an adapted organism. Even in the early stages of embryogenesis certain groups of cells are designed to grow into a certain organ, and this specifies their functionality at all levels, including the molecular level.

An organism is similar to a machine built according to some design for the achievement of definite purposes. Moreover, it is a machine of a high level of complexity, capable of responding rationally to events that have not been planned in advance. An example is the fact that the organism of a higher animal manufactures antibodies in response to practically any antigenes, including artificial ones, which can never be encountered in nature. The finalistic description is directly based on the historical evolution of living organisms. This is not so in ordinary physics and chemistry. The meaninglessness of the following statement, for example, is obvious: "The sodium and chloride ions interact with each other in order to build a cubic crystal." On the contrary, the statement "... since the Na^+ and Cl^- ions have single charges and such-and-such radii, an ionic crystal of NaCl must be cubic" is meaningful. The biologists often ask "what for?" and the physicists ask "why?". Evidently, it is the second question that has a true scientific meaning. Physics and natural sciences in general are causal—science is seeking the causes of phenomena.

As a matter of fact, there is no contradiction between finalism and causality and the above difference between physics and biology is only superficial. Finalism arises in physics every time when we

deal with stability problems, with variational principles. We mean stability in a strict mathematical sense, according to Lyapounov (see Chapter 15). The stable state of a dynamical system is preserved at small perturbations—after being deviated from this state the system returns to it. The finalistic formulation is as follows: the system tends to preserve its state. On the contrary, the unstable state is changed irreversibly at small perturbations—the system tends to pass over to another state. As an example, we may cite the stable and unstable states of equilibrium of a simple physical pendulum.

The variational principle is always finalistic. For instance, according to Hamilton's principle of least action, the variation of action is equal to zero, the action is at a minimum. "The purpose of a mechanical system consists in its least action." But, as shown by classical mechanics, Hamilton's principle is equivalent to the Lagrangian equations of motion, which in turn follow from Newton's second law. This law is causal: it describes an accelerated motion as a result of the action of forces. Other examples of finalistically formulated laws of physics are: Fermat's principle in optics, Le Châtelier's principle in thermodynamics, Lenz's law in electrodynamics. The variational finalism reduces to causality. The number of such examples is unlimited.

1.6. The Properties of Open Systems

As has already been said, living systems are basically open and are therefore nonequilibrium. One of the first to understand this was Bauer, a Soviet biologist, who wrote that "... the nonequilibrium state of living matter and, hence, its constantly maintained activity is due to ... the molecular structure of living matter, and the source of the work performed by living systems is in the long run the free energy inherent in this molecular structure, in this state of molecules." The subsequent development of the thermodynamics of open systems as applied to biology is associated with the names of Bertalanffy and, mainly, of Prigogine and his colleagues.

A living organism is a thermodynamically open system which undergoes chemical reactions. Biochemical reactions are catalytic at all stages and protein-enzymes serve as catalysts. The change in entropy of such a system is expressed by the sum of the entropy produced inside the system, d_iS, and the entropy supplied externally or given off to the surroundings, d_eS:

$$dS = d_iS + d_eS \tag{1.1}$$

The quantity d_iS is always positive according to the second law of thermodynamics—if the organism is placed in an insulating shell, the entropy flow d_eS will be equal to zero and the entropy can only

increase. The quantity d_iS defines the entropy production as a result of chemical reactions taking place inside the system. The sign of d_eS depends on the concrete situation. Equation (1.1) shows the possibility of a steady nonequilibrium state of an open system. In the stationary state the thermodynamic quantities that characterize the system are constant but they have no equilibrium values. The entropy of the system is not at a maximum. For the *stationary state* we have:

$$dS = 0, \text{ i.e., } d_eS = -d_iS < 0 \tag{1.2}$$

In other words, the entropy produced is given up completely to the surroundings.

Let us consider an isolated system consisting of a living organism and its environment. The organism receives food, oxygen, water, and in turn gives off various substances to the surroundings. The organism and its surroundings exchange heat. An astronaut in a spaceship is practically under such conditions. The astronaut's organism is an open system with respect to the spaceship, which is well insulated against the surrounding outer space. According to the second law of thermodynamics, the total entropy change of the entire system is given by

$$dS = dS_1 + dS_2 > 0 \tag{1.3}$$

where dS_1 and dS_2 are the entropy changes of the astronaut and the surroundings. We have then:

$$dS_1 = d_iS_1 + d_eS_1 \tag{1.4}$$

The entropy change dS_2 occurs only as a result of the exchange of matter and energy between the astronaut and the environment; no entropy is practically produced in the environment that surrounds the astronaut. Hence,

$$dS_2 = -d_eS_1 \text{ and } dS = d_iS_1 > 0 \tag{1.5}$$

If the state of the astronaut is stationary, then

$$dS_1 = 0, \; d_eS_1 = -d_iS_1 < 0, \; dS_2 = dS = d_iS_1 > 0 \tag{1.6}$$

Thus, the astronaut's steady state is maintained by an increase in entropy in the surroundings determined by the entropy flow from the astronaut's organism to the environment which compensates for the entropy production in the organism. In this lies the meaning of Schrödinger's words: "the organism feeds on negative entropy". The value of dS_2 is greater than zero because of the loss of heat by the astronaut and also because the entropy of the substances given off by the astronaut is higher than the entropy of the substances consumed by him. The stationary state of the astronaut will be maintained until the nutrients in the surroundings are exhausted or the irrevers-

ible processes taking place in the astronaut's organism lead to its change (ageing). The steady state may be long but not infinite. Its attainment is determined by the presence of two time scales—a rapid scale for the time of the exchange of entropy with the environment and a relatively slow scale for the time of exhaustion of the reserves of nutrients and/or of the ageing of the organism.

Let us consider one more example. Suppose we have two heat reservoirs, one at temperature T_1 and the other at temperature T_2; the two reservoirs are connected with each other by a thin heat conductor. A constant heat flow is set up in the conductor and the entire length of this conductor is in the stationary state. This state is maintained until the temperatures T_1 and T_2 are changed to any noticeable extent; these temperatures are gradually equilibrated. The time of attainment of a stationary state is much shorter than the time of attainment of the final equilibrium at which $T_1 = T_2$. In calculating the steady-state conditions on the fast time scale the temperatures T_1 and T_2 are considered to be constant—they undergo change on the slow scale.

The existence of the biosphere itself may be regarded as a steady-state process accomplished against the background of the enormous irreversible process of the Sun's cooling.

The stationary state of an open system may be close to or far from equilibrium, it may be stable or unstable. The nonstationary state in turn may be close to or far from the stationary state. As we shall see (see Chapter 9), the thermodynamics of open systems that are close to equilibrium, is different from the thermodynamics of far-from-equilibrium open systems. Here strict quantitative criteria of nearness or remoteness are introduced. A number of facts indicate that biological systems are far from equilibrium. Biological evolution is possible only in a system that is far from equilibrium.

From what has been said above it does not follow, however, that the application of reversible (equilibrium) thermodynamics to any biological phenomena is impossible. Dynamical, irreversible life processes occur inside an organized structure, which changes slowly or remains practically unaltered. This holds for all levels of biological structure, beginning with the molecular level. Accordingly, in structural studies we may use the conditional thermodynamic equilibrium. The origin of a biological structure itself may be roughly visualized as consisting of two stages—the biosynthesis of the constituent elements (macromolecules, cells) and the assembly of an organized structure from these elements. The process of assembly is to a considerable extent thermodynamically controlled, say, at the molecular level the system tends to a state of lowest chemical potential. Reversible thermodynamics proves to be at the basis of molecular biophysics.

At the same time it is obvious that in biology we encounter a complex set of thermodynamics and kinetic phenomena. The difference between thermodynamics and kinetics is illustrated by the progress of a chemical reaction. A chemical reaction is basically possible only if it is accompanied by a decrease in the free energy, $\Delta G < 0$. It is a necessary but not a sufficient condition for the reaction to take place. If the initial and final states of the system are separated by a high potential barrier, the rate constant of the reaction, which depends exponentially on the height of the energy barrier E_a according to the Arrhenius law

$$k = A \exp(-E_a/RT) \tag{1.7}$$

(where A is the pre-exponential factor, R is the gas constant, and T is the absolute temperature) may be vanishingly small. Therefore, we may regard the nonequilibrium state protected by a high barrier as an equilibrium state. An example is the oxyhydrogen gas, a mixture of H_2 and O_2, which detonates upon activation of a chain reaction by the flame of a match and which is stable, i.e., in a quasiequilibrium state in the absence of activation. The role of the catalyst consists in lowering the activation barrier E_a.

Irreversible (nonequilibrium) thermodynamics proves to be necessary for the treatment of dynamic processes, primarily the evolution phenomena.

Directly associated with thermodynamics is the theory of information. A change in the amount of information about a particular system which is regarded as a message is proportional to the entropy change (see Chapter 9). The informational aspects of biophysics are of considerable interest. There is revealed the heuristic limitedness of the significance of the amount of information (and, hence, of the entropy) in treatments of developing biological systems. It becomes necessary to consider the reception of information and the creation of new information. Either of these is possible only under nonequilibrium, nonsteady-state and instability conditions. It is not the amount of information that is important in biology but its quality, meaning or value (Chapter 17).

1.7. Scope and Methods of Biophysics

Living organisms are multi-level complex systems. Large and small molecules, cell organelles, cells, tissues, organs, organisms, populations, biocenosis, the biosphere are all levels that must be dealt with in biology and biophysics.

Biophysics may be subdivided into three branches: molecular biophysics, biophysics of the cell, and biophysics of complex systems. This division is not obligatory but it is convenient. Let us

characterize the contents and methods of the three branches of biophysics.

Molecular biophysics is concerned with the study of the structure and physico-chemical properties of biologically functional molecules, primarily biopolymers—proteins and nucleic acids. The objective of molecular biophysics is to disclose the physical mechanisms responsible for the biological functionality of molecules, say, for the catalytic activity of protein-enzymes. Molecular biophysics is the most developed branch of biophysics. It is inseparable from molecular biology and chemistry. The great advances made in this area are understandable—it is easier to study molecules (even protein molecules, the most complex of those known to science) than cells or organisms. Molecular biophysics rests, on the one hand, on biochemical disciplines (biochemistry, molecular biology, bioorganic and bioinorganic chemistry) and, on the other, on the physics of small and large molecules. Accordingly, we shall first consider the chemical foundations of biophysics (Chapter 2) and macromolecular physics (Chapter 3), and only after this shall we turn to molecular biophysics as such (Chapter 4-8).

Since molecular biophysics is mainly concerned with the structure of molecules and their functionality, we may consider the equilibrium properties of molecules. The theoretical apparatus of molecular biophysics consists of equilibrium (reversible) thermodynamics, statistical mechanics, and, of course, quantum mechanics, since what we are concerned with is chemistry, molecules. For an experimental study of biologically functional molecules (we shall call them simply biological molecules or biomolecules) use is made of a wide range of physical methods. These are, first of all, the methods employed in macromolecular physics for the determination of molecular masses, sizes, and shapes—sedimentation in the ultracentrifuge, scattering of light and X-rays by solutions of the substances to be studied, etc. The second group includes methods used for the investigation of molecular structure and based on the interaction of matter with light in the broad sense of these words, from X-rays to radiofrequency radiation. These are optical and spectroscopic techniques which include X-ray diffraction analysis, gamma-resonance spectroscopy (the Mössbauer effect), electronic and vibrational spectra, i.e., absorption and luminescence spectra in the ultraviolet and visible regions, infrared and Raman spectra. Here also belongs spectropolarimetry, i.e., the investigation of the natural and magnetic rotation of the plane of polarization of light and of circular dichroism. Very valuable information is furnished by nuclear magnetic resonance and electron paramagnetic resonance (NMR and EPR) spectra. In the case of EPR the application of paramagnetic spin labels is especially important. The third group includes calorimetric methods used for the study of the conversions of biological macro-

molecules. And, finally, the structure of proteins and nucleic acids (and also, of course, supermolecular structures, cell organelles and cells) is directly studied by means of electron microscopy.

The application of a number of these methods in biology is specific. It is therefore necessary to dwell on these methods in a course of biophysics. Chapter 5 is devoted to a brief description of the theory of optical and spectroscopic methods and of the biophysical results obtained by means of these methods.

Molecular biophysics naturally unites with the biophysics of the cell, which deals with the structure and functionality of cell and tissue systems. This branch of biophysics is the oldest and traditional one. Its main concern today is the study of the physics of biological membranes and bioenergetic processes. The biophysics of the cell includes the study of the generation and transmission of the nerve impulse, mechanochemical processes (in particular, muscular contraction), photobiological phenomena (photosynthesis, vision). The experimental techniques listed above are also used for these purposes. The biophysics of the cell deals with more complex problems and is confronted with more serious difficulties than molecular biophysics. The cell biophysics is described in Chapters 10-14.

The biophysics of complex systems is predominantly a theoretical branch of biophysics devoted to general physico-biological problems and physico-mathematical modelling of biological processes. The basic modern sections of the theoretical biophysics of complex systems are as follows:

1. The general theory of dissipative nonlinear dynamical systems—the thermodynamics of irreversible processes and kinetic modelling.
2. The theory of excitable media, which includes the theory of biological oscillatory processes.
3. The general theoretical treatment of bioenergetic phenomena.
4. The general theory and modelling of the processes of biological development, including various approaches to these phenomena based on information theory.

These areas of biophysics are dealt with in Chapters 9, 15, 16, and 17; we shall also encounter the same general ideas in chapters concerned with problems of the biophysics of the cell.

Chapter 2 Chemical Foundations of Biophysics

2.1. Chemistry and Biology

The living cell and living organisms are complex chemical machines. They exist owing to the chemical transformations of substances that are supplied from the outside world and to the flow of the substances into the surroundings due to metabolism. Biology and biophysics are inseparably tied up with chemistry.

No physical methods alone would have been able to establish the structure of complex molecules without chemical investigations. Chemistry solves these problems by studying the structure of substances and their transformations in chemical reactions.

The chemistry of life, or organic chemistry, was originally remote from inorganic chemistry. It was believed to be a reliable support of vitalism until scientists came to learn how to synthesize organic substances from nonliving matter (the first experiment was the synthesis of urea, $CO(NH_2)_2$, by Friedrich Wöhler in 1828). Later, organic chemistry ceased to be the chemistry of living matter and became the synthetic chemistry of carbon compounds, the chemistry of hydrocarbons and their derivatives. An almost independent path was taken up by biochemistry, a science concerned with the structure and properties of biological molecules, with the progress of chemical reactions in living organisms. Biochemistry has made enormous advances in unravelling of complicated metabolic pathways. The union of biochemistry and physics gave rise to *molecular biology*, which deals with the physico-chemical, molecular interpretation of basic biological phenomena, primarily heredity. At the same time, organic chemistry again turned to living matter on the basis of the long experience of investigations of organic compounds. Thus arose *bioorganic chemistry* and later *bioinorganic chemistry*, whose concern is the study of biological molecules containing metal atoms. It is impossible to draw a boundary line between the above-mentioned areas of investigations of the chemistry of life and there is no need for doing this.

Let us note the main specific features of the chemistry of living matter—the chemistry of biomolecules.

1. A living system is necessarily chemically heterogeneous. It is pointless to speak of living molecules—biological molecules taken separately cannot live, i.e., they do not differ from any other inanimate molecules.

2. Living matter is characterized by the unity of chemical structure. The enormous diversity of biological species and individuals does not imply the extreme variety of biological molecules and biochemical reactions. The main substances and chemical mechanisms are the same in all living things. All proteins are built up of 20 amino-acid residues, and all nucleic acids, of 4 nucleotides. The same atomic structures appear in all living organisms. The fundamental biochemical processes are also the same in all living organisms. The variety of organisms is due to the large variety of combinations of the same atomic groupings and their interactions.

3. The structure and properties of living cells and organisms are dictated, in the final analysis, by the nucleic acids (DNA and RNA) which have a "legislative" power in the sense that they specify the genetic program of protein synthesis. In their turn, proteins function as "executives" for the simple reason that no single chemical reaction in the cell proceeds without the participation of a specific enzyme.

4. Biochemical processes and biological molecules are the result of the evolutionary development. The biological evolution was preceded by a chemical evolution (Chapter 17) and later these became tied up inseparably with each other. The species and organisms are characterized by the biochemical, molecular adaptation to environmental conditions. In studying the chemistry of living matter one must always consider the biological development.

5. Life is characterized by a precise and subtle individualization. The minor differences in the structure of molecules, say, the differences between the methyl and ethyl groups, which are of no substantial importance in ordinary chemistry, are very important in biology (ethyl alcohol, C_2H_5OH, causes intoxication, and methyl alcohol, CH_3OH, causes blindness). Biological molecules and macromolecules have a strictly definite composition and structure.

6. Chemical reactions that take place in living organisms are strictly controlled by forward and reverse pathways in multistage metabolic processes as well as by the spatial separation of reactions made possible due to *compartmentation* realized by cells and intracellular membranes. Delicate concentration gradients are maintained.

7. Finally, biological molecules are built up of atoms of light elements: C, H, O, N, P, S. Besides, the functionality of the ions of alkali and alkaline-earth metals (Na, K, Ca, Mg) is universal in organisms. An important, though not universal, role is played by small amounts of other metals (Fe, Zn, etc., up to Mo inclusive).

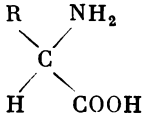
The human body contains (in atomic percent) hydrogen (60.56), oxygen (25.67), carbon (10.68), nitrogen (2.44), calcium (0.23), phosphorus (0.13), sulphur (0.13), sodium (0.08), potassium (0.04), chlorine (0.03), magnesium (0.01), and much smaller amounts of Fe, Zn, Cu, Co, Se, F, and I. The excess amounts of H and O present are due to a large content of water.

Further in this chapter we shall briefly characterize the most important biological molecules and biochemical processes. Our task is not to duplicate courses of biochemistry but rather to summarize basic information necessary for expounding a course of biophysics.

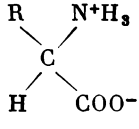
2.2. Amino Acids

All proteins are built up of 20 types of amino-acid residues; a number of amino acids that are not contained in proteins participate in metabolism. Evidently, the amino acids originated on the earth at the first stage of chemical evolution about 4×10^9 years ago, as a result of chemical reactions in the primordial reducing atmosphere and in the ocean upon absorption of energy (see Chapter 17). The simplest amino acids are also encountered in meteorites.

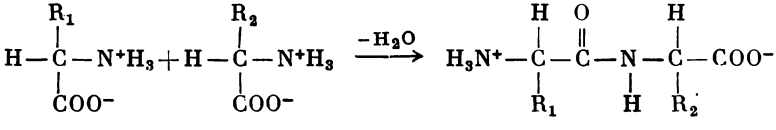
Proteins contain the residues of α-amino acids, which have the structure



where R is a radical [a hydrocarbon radical or a radical that contains, apart from C and H, other atoms too (O, S, N)]. As has been shown by a study of the electrical properties of amino acids (see Sec. 2.3), their structure is expressed more exactly by the formula of a *dipolar ion* (also called *zwitterion*):

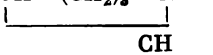
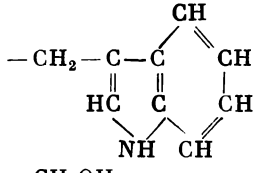
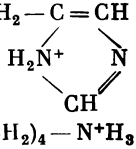
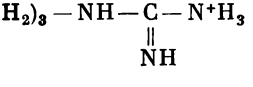


The formation of a protein or polypeptide chain occurs by way of the polycondensation of amino acids:



Here peptide bonds, —NH—CO—, are formed.

Table 2.1. The Common Amino Acids

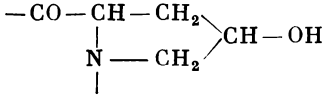
Name of amino acid	Abbreviation (three-letter symbol)	Abbreviation (one-letter symbol)	R group
I. *Neutral Amino Acids*			
1. Glycine	Gly	G	$-H$
2. Alanine	Ala	A	$-CH_3$
3. Valine	Val	V	$-CH(CH_3)_2$
4. Leucine	Leu	L	$-CH_2-CH(CH_3)_2$
5. Isoleucine	Ile	I	$-CH(C_2H_5)CH_3$
6. Phenylalanine	Phe	F	$-CH_2-C_6H_5$
7. Proline*	Pro	P	
8. Tryptophan	Trp	W	
9. Serine	Ser	S	$-CH_2OH$
10. Threonine	Thr	T	$-CH(OH)CH_3$
11. Methionine	Met	M	$-CH_2-CH_2-S-CH_3$
12. Asparagine	Asn	N	$-CH_2-CO-NH_2$
13. Glutamine	Gln	Q	$-CH_2-CH_2-CO-NH_2$
14. Cysteine**	Cys	C	$-CH_2SH$
II. *Acidic amino acids (in the form of anions)*			
15. Aspartic acid	Asp	D	$-CH_2-COO^-$
16. Glutamic acid	Glu	E	$-CH_2-CH_2-COO^-$
17. Tyrosine	Tyr	Y	$-CH_2-C_6H_4-O^-$
III. *Basic amino acids (in the form of cations)*			
18. Histidine	His	H	
19. Lysine	Lys	K	$-(CH_2)_4-N^+H_3$
20. Arginine	Arg	R	

* Proline is not an amino acid; it is an imino acid which contains an imino group, NH, instead of the amino group NH_2.

** As a rule, two cysteine residues are joined together in proteins by a disulphide linkage with loss of the hydrogen atom ($-CH_2-S-S-CH_2-$). The amino-acid residue formed is called *cystine* and is designated Cys-Cys or $(Cys)_2$ or CySSCy (the abbreviated one-letter symbol is also C).

Table 2.1 gives a list of 20 amino-acids (—CO—CHR—NH—). The amino-acid residues listed in Table 2.1 are classified according to electrochemical properties. Among them are aliphatic amino acids (Ala, Val, Gly, Ile, Leu), hydroxyl-containing amino acids (Ser, Thr), amino acids containing an amino group (Asn, Gln), amino acids containing a carboxyl group (Asp, Glu), sulphur-containing amino acids (Cys, Met), and amino acids containing aromatic π-electronic rings (Tyr, Phe, Trp, His). These amino-acid units are responsible for all the variety of the properties and structure of proteins.

Apart from the 20 common amino acids, proteins also contain the derivatives of certain groups. Important among these minor residues is hydroxyproline



since it is found in a considerable amount in one of the universal animal proteins, collagen.

2.3. Electrolytes

Living organisms contain numerous ions—small organic and inorganic cations and anions, basic and acidic groups of amino-acid residues in proteins and of nucleotides in nucleic acids. The dissociation of electrolytes into ions occurs due to the presence of a water environment. The theory of electrolytes is the concern of physical chemistry courses; here we shall confine ourselves to brief information needed for our subsequent discussion.

According to Brönsted, an acid is a molecule that can give up a proton and a base is a molecule that can accept a proton. Water is an acid in the reaction $H_2O \rightleftarrows H^+ + OH^-$ and a base in the reaction $H_2O + H^+ \rightleftarrows H_3O^+$ (the hydronium ion).

Water is a very weak electrolyte and its dissociation constant K is low. At 25°C in water,

$$[H^+][OH^-] = K[H_2O] = 10^{-14} \text{ mole}^2/\text{litre}^2 \qquad (2.1)$$

Since $[H^+] = [OH^-]$, the hydrogen-ion concentration in water is 10^{-7} mole/litre. Instead of the hydrogen-ion concentration, it is more convenient to use its negative logarithm denoted by the symbol pH. For water, i.e., for a neutral medium, pH = 7. Accordingly, for acids pH < 7 and for bases pH > 7.

Biological electrolytes are usually weak, i.e., they dissociate slightly. Let us consider the dissociation of an acid, $HA \rightleftarrows H^+ + A^-$. The dissociation constant, i.e., the equilibrium constant K,

of this reaction is

$$K = \frac{a_H a_A}{a_{HA}} \tag{2.2}$$

where $a_H = a_A$ and a_{HA} are the activities of the corresponding substances in solution. In solutions approaching ideal solutions, the activities are practically equal to molar concentrations. Therefore,

$$[H^+]^2 = K[HA] \tag{2.3}$$

and, introducing the notation $pK = -\log K$, we have

$$pH = \frac{1}{2}(pK - \log [HA]) \tag{2.4}$$

The ionization constant of a weak acid in water is given by

$$\alpha = \frac{[A^-]}{[A^-]+[HA]} \approx \frac{[H^+]}{[HA]} = \sqrt{\frac{K}{[HA]}} \tag{2.5}$$

The degree of dissociation or the strength of an acid increases with rise of K, i.e., it falls off with increasing pK; the strength of a base declines with increase of K, i.e., increases with increasing pK. The condition $[A^-] = [H^+]$ holds only for a neutral solution. Generally speaking,

$$pH = pK \pm \log \frac{a_A}{a_{HA}} \tag{2.6}$$

or (in an ideal solution)

$$pH = pK \pm \log \frac{\alpha}{1-\alpha} \tag{2.7}$$

The plus sign refers to the acid and the minus sign to the base, for which

$$\alpha = \frac{[HA]}{[A^-]+[HA]} \tag{2.8}$$

which corresponds to the reverse reaction $A^- + H^+ \rightleftarrows HA$, where A^- is a base.

In a nonideal solution, the activities differ from the concentrations. We have

$$pH = pK + \log \frac{\gamma_A}{\gamma_{HA}} \pm \log \frac{\alpha}{1-\alpha} \tag{2.9}$$

where γ_A and γ_{HA} are the activity coefficients.

The nonideality of an electrolyte solution is determined by the interaction of ions with each other and with solvent molecules. At a sufficiently high concentration of ions, each of them is surrounded by an *ionic atmosphere*—by ions of opposite charge (counterions).

The *Debye-Hückel theory*, which is based on classical statistical mechanics and electrostatics, allows one to determine the free energy

of interionic interactions in solution. It deals with a system consisting of a central ion and an ionic atmosphere with a net charge opposite in sign to the central ion. The free energy of the interaction per unit volume of solution is given by

$$G = -\frac{\kappa^3 kT}{12\pi} \tag{2.10}$$

where k is Boltzmann' constant, T is the temperature, and κ is the Debye-Hückel parameter, which is related to the effective radius of the ion atmosphere and which has the dimensions of reciprocal length. It is sometimes referred to as the Debye-Hückel reciprocal length. The ion-atmosphere radius κ is expressed in terms of ionic charges e_i and the number of ions in unit volume, n_i:

$$\kappa^2 = \frac{4\pi}{\varepsilon kT} \sum_i e_i^2 n_i \tag{2.11}$$

The quantity ε is the dielectric constant of the solvent; the summation is carried out over all the ionic species i. The activity coefficient γ_i of a given species of ions is found from the change in free energy upon removal of the ion from the solution:

$$\frac{\partial G}{\partial n_i} = -\frac{e_i^2 \kappa}{2\varepsilon} = -\frac{|e|^2 z_i^2 \kappa}{2\varepsilon} = kT \ln \gamma_i \tag{2.12}$$

where $|e|$ is the absolute value of the electronic charge and z_i is the valence of the ion. We have

$$\ln \gamma_i = -\frac{1}{2} \frac{e^2 z_i^2 \kappa}{\varepsilon kT} = -\frac{|e|^3 z_i^2 (2\pi)^{1/2}}{(\varepsilon kT)^{3/2}} I^{1/2} \tag{2.13}$$

where I is called the *ionic strength* and is defined by

$$I = \frac{1}{2} \sum_i n_i z_i^2 \tag{2.14}$$

The ionic strength I has the dimensions of reciprocal volume. The ionic strength characterizes the total number of ionic charges in unit volume, irrespective of their sign. The ionic strength and the quantities pH and pK are the most important characteristics of an ionic solution.

Thus, the activity coefficients of ions of a given species depend on the concentration of ions of all the species present.

The Debye-Hückel theory is valid for strong electrolytes. The amino acids are weak electrolytes. However, at high ionic strengths the activity coefficients of the amino acids are different from unity. At low ionic strengths $\gamma_A \approx \gamma_{HA} \approx 1$ and [cf. Eq. (2.9)]

$$\mathrm{pH} \approx \mathrm{pK} \pm \log \frac{\alpha}{1-\alpha} \tag{2.15}$$

The titration of a weak acid by a strong base, say NaOH, involves the reaction

$$HA + Na^+ + OH^- \rightarrow A^- + Na^+ + H_2O$$

Hence, $[A^-] \approx [Na^+]$. The degree of ionization α is equal to the ratio of the concentrations of strong and weak electrolytes:

$$\alpha = \frac{[A^-]}{[A^-] + [HA]} \approx \frac{[Na^+]}{[HA]} \tag{2.16}$$

The determination of α, pH, and, hence, of pK is carried out by way of the titration of an acid with a base and of a base with an acid. For crude determinations, use is made of chemical indicators; for precise determinations, potentiometric methods are employed which are based on the recording of the potential on the electrode immersed in the solution. The potential undergoes change upon titration. To determine pH a glass electrode is used. This electrode and a comparison electrode (usually the calomel electrode) are placed in a solution of known concentration of the substance under study. The potential difference is measured directly in pH units with suitable calibration.

The amino acids are amphoteric electrolytes characterized by two values of pK which correspond to the titration of acidic groups (COO^-) with a base (pK_1) and to the titration of basic groups (N^+H_3) with an acid (pK_2). Table 2.2 gives some values of pK_1, pK_2, and pK_3 (for amino acids with ionogenic radicals) and also of pH_i. The values of pH_i correspond to the *isoelectric point*. The amino acid moves to the cathode if it is charged positively and to the anode if charged negatively. At the isoelectric point the molecule of an amphoteric electrolyte has a balance of positive and negative charges (i.e., it is neutral) and does not move in an electric field. For neutral amino acids we have

$$pH_i = \frac{1}{2}(pK_1 + pK_2) \tag{2.17}$$

The study of the electrochemical properties of amino acids proves their dipolar structure (see page 27). The heats of reactions of ionization of organic acids

$$RCOOH \rightleftharpoons RCOO^- + H^+$$

are low, being of the order of 4 kJ/mole. For the dissociation of substituted ammonium ions

$$RN\ H_3 \rightleftharpoons RNH_2 + H^+$$

the reaction heats are high, of the order of 50 kJ/mole. The heats of ionization for amino acids range from 5.5 to 8.8 kJ/mole in acid-

Table 2.2. The Electrochemical Constants of Amino Acids

Amino acid	pK_1	pK_2	pK_3	pH_i
Glycine	2.35	9.78		6.1
Alanine	2.34	9.87		6.1
Valine	2.32	9.62		6.0
Leucine	2.36	9.60		6.0
Serine	2.21	9.15		5.7
Proline	1.99	10.60		6.3
Tryptophan	2.38	9.39		5.9
Aspartic acid	2.09	3.87 (COO^-)	9.82 (NH_3^+)	3.0
Glutamic acid	2.19	4.28 (COO^-)	9.66 (NH_3^+)	3.2
Tyrosine	2.20	9.11 (NH_3^+)	10.1 (O^-)	5.7
Cysteine	1.96	8.18 (NH_3^+)	10.28 (S^-)	5.1
Arginine	2.02	9.04 (NH_3^+)	12.48 (guanidine)	10.8
Lysine	2.18	8.95 (NH_3^+)	10.53 (NH_3^+)	9.7
Histidine	1.77	6.10 (imidazole)	9.18 (NH_3^+)	7.6

ic solution and from 42 to 55 kJ/mole in basic solution. Hence, the following reactions take place:

$$H_3N^+-CHR-COO^- + H^+ \rightleftharpoons H_3N^+-CHR-COOH$$
$$H_3N^+-CHR-COO^- + OH^- \rightleftharpoons H_2N-CHR-COO^- + H_2O$$

The dipolar-ion structure of amino acids is also evidenced by a strong increase of ε upon their dissolution in water, high densities, and high melting points of solid amino acids, which is determined by a strong electrostatic interaction.

While working with biological substances one should maintain the pH values constant. Stabilization of pH is achieved by means of *buffer solutions*. In the presence of neutral salts, the dissociation of weak acids and bases does not depend on dilution. Let us consider a solution of a weak acid HA and its sodium salt NaA. The equilibrium constant of the reaction

$$HA \rightleftharpoons H^+ + A^-$$

is equal to

$$K = \frac{[H^+][A^-]}{[HA]} \tag{2.18}$$

and since $[A^-] = [H^+]$

$$[H^+] = \sqrt{K[HA]} \tag{2.19}$$

In the presence of the salt NaA, which has been dissociated much more strongly than the acid, $[A^-] \approx [NaA]$. Thus,

$$K = \frac{[H^+][NaA]}{[HA]} \tag{2.20}$$

and

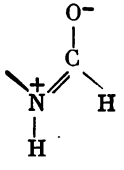
$$[H^+] = K\frac{[HA]}{[NaA]} \tag{2.21}$$

The hydrogen-ion concentration, $[H^+]$, and, hence, pH depend on the ratio of the acid and salt concentrations but are independent of the extent of dilution. Thus, in a 0.1 *N* solution of acetic acid and a 0.1 *N* solution of sodium acetate in water the value of pH is 4.628. With a ten-fold dilution of the solution pH = 4.670, with a hundred-fold dilution the value of pH is 4.73. When 1 cm^3 of 0.01 *N* HCl is added to 1 litre of water, the pH value decreases from 7 to 5. When the same amount of HCl is added to the solution indicated the pH value falls from 4.628 to 4.540. Apart from acetate buffers, use is also made of phosphate and other buffer solutions.

2.4. The Composition and Primary Structure of Proteins

The macromolecules of proteins are composed of one or more polypeptide chains made up of amino-acid residues. One end of the chain bears a free amino group NH_2 (the so-called N-terminal end) and the other end carries a free carboxyl group COOH (the C-terminal end). The characteristic molecular masses of individual polypeptide chains in proteins are of the order of 20,000, which corresponds to 150-180 amino-acid residues (the average molecular mass of an amino-acid residue is 120). Molecules that contain less than 100 residues are usually called polypeptides rather than proteins. Some hormones are such polypeptides.

The peptide bond —CO—NH—, which joins together amino acids in proteins, has a specific planar structure, as has been found by X-ray structure analysis (Pauling and Corey). All the four atoms in the peptide bond lie in one plane (Fig. 2.1). The N—C bond is shorter than the same bond in aliphatic amines $R—NH_2$, where its length is equal to 0.147 nm. This shortened length of the N—C bond, as well as the planar arrangement of the bonds, is evidence of the conjugation of the N—C and C = O bonds and, hence, of the overlap of their electron shells, this being accompanied by a shift of electron density from N to C. This can be shown by the contribution of the structure



that is, the N—C bond is partially double and the C=O bond is partially single.

The amino-acid composition of a protein is now established by means of automatic devices. A protein is broken down into amino acids by *hydrolysis*—the reverse of the polycondensation reaction:

$$-CHR_1-NH-CO-CHR_2-+H_2O \rightarrow -CHR_1-NH_2+HOOC-CHR_2-$$

The hydrolysis occurs when the protein is acted on by bases, acids, and also by proteolytic enzymes (*proteases*) which catalyze the

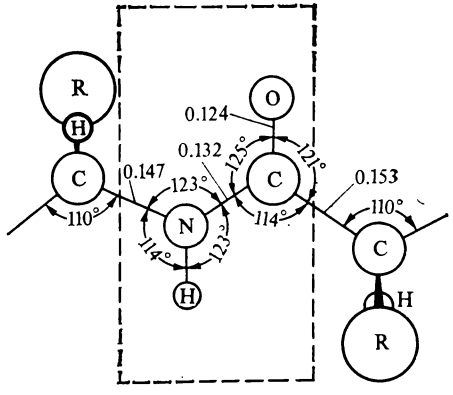


Fig. 2.1. The peptide bond.

rupture of peptide bonds. The resulting hydrolysate, a solution of a mixture of amino acids, is analysed by *chromatographic techniques*. For example, proteolytic hydrolysis occurs during digestion—the food proteins are cleaved in the digestive tract into the constituent amino acids, and then proteins demanded by the body are synthesized from these amino acids.

The composition of proteins is not uniform—different amino acids occur in proteins with different frequencies. There are rare and frequent amino-acid residues, just as there are seldom and frequently encountered letters in a text. The most seldom encountered residues are Trp, Met, and His, and the most frequently encountered ones are Ala, Ser, and Gly. This is valid on the average—some specialized proteins have a specific composition, which is different from the average composition (for example, collagen, see Sec. 4.9).

The amino-acid sequence in the polypeptide chain(s) of a protein is called the *primary structure*. The primary structure is determined through partial hydrolysis of a protein by means of proteases, which catalyze the cleavage of the peptide linkage only between definite amino-acid residues. For example, trypsin hydrolyzes only those peptide bonds which are formed by the CO groups of the basic amino

acids arginine and lysine. This results in the formation of a mixture of peptides—short fragments of the protein chain. The identification of these fragments is made by means of chemical and physico-chemical methods (chromatography, mass spectroscopy). Using another enzyme, one can hydrolyze other bonds in the protein and obtain a mixture of other fragments (peptides).

Knowing the structure of peptides produced by hydrolysis with different enzymes, it is possible to establish the primary structure of the protein, solving crossword-like puzzles. Today, primary structures are known for many hundreds of proteins. The primary structures of proteins are regularly published in the atlas of protein structures. Figures 2.2 and 2.3 show the primary structures of bovine ribonuclease and sperm whale myoglobin. In the first case there are four disulphide linkages, which join the cysteine residues together.

The *primary structure of a protein* is a kind of text written by means of a twenty-letter alphabet. The sense and content of this text consists in the biological functioning of the protein which is determined, in the final analysis, by the primary structure. The protein texts reflect the biological evolution—the comparison of homologous proteins, which perform the same function in different species, allows one to disclose the differences in the texts. These differences determined by the mutational amino-acid replacements are the greater, the farther apart are the biological species. Especially thoroughly studied in this respect is cytochrome *c*, the protein of the respiratory chain which is universal to all living organisms. Comparison of the primary structures of cytochromes *c* of different kinds permits one to construct the evolution tree.

The primary structure of a protein can also vary as a result of mutations. There appear "errors" in the protein text which have, as a rule, a negative effect on the functions of the protein. A number of hereditary blood diseases are associated with the mutations of hemoglobin. Sickle-cell anemia, a severe inherited disease, is caused by the substitution of valine for glutamic acid at the sixth position in the β-chains of hemoglobin. The man's hemoglobin is made up of four chains—two identical α-chains, each of which contains 141 amino-acid units, and two identical β-chains containing 146 amino acids each. Thus, the substitution of even two amino-acid residues out of 574 units leads to very serious consequences. In this case, the nonpolar neutral amino-acid residues are replaced by acidic units. At present, over 100 human mutant hemoglobins are known to exist.

2.5. Nucleic Acids

The second, most important class of biopolymers includes nucleic acids, the macromolecules that are responsible for the biosynthesis of proteins, for the assembly of their primary structures.

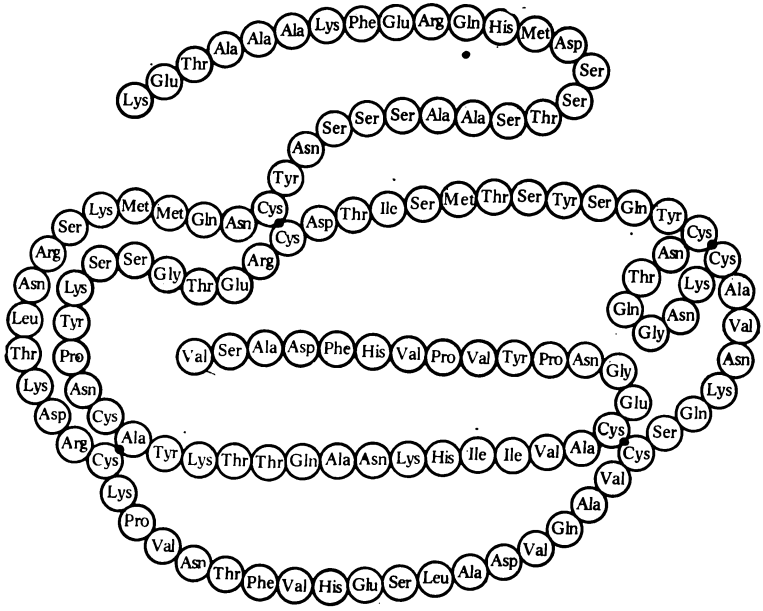

Lys Glu Thr Ala Ala Ala Lys Phe Glu Arg Gln His Met Asp Ser Ser Ser Ser Ala Ala Ser Thr Ser Ser Asn Tyr Cys Lys Met Met Gln Asn Cys Asp Thr Ile Ser Met Thr Ser Tyr Ser Gln Tyr Cys Ser Cys Asn Cys Arg Ser Ser Gly Thr Glu Arg Thr Lys Ala Asn Lys Gln Asn Val Leu Tyr Gly Asn Thr Pro Val Ser Ala Asp Phe His Val Pro Val Tyr Pro Asn Gly Asn Lys Asp Cys Glu Lys Gln Arg Cys Ala Cys Ser Tyr Lys Thr Thr Gln Ala Asn Lys His Ile Ile Val Ala Cys Val Lys Pro Val Asn Thr Phe Val His Glu Ser Leu Ala Asp Val Gln Ala

Fig. 2.2. The primary structure of bovine ribonuclease.

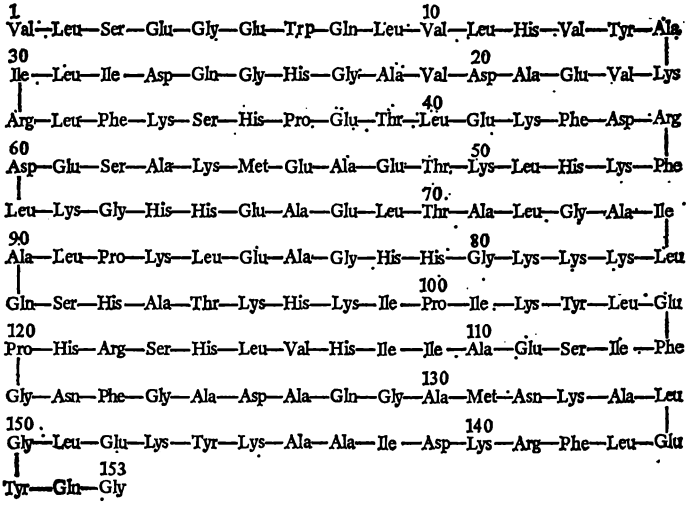
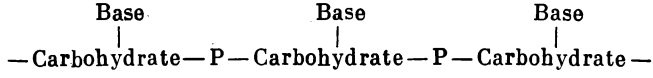
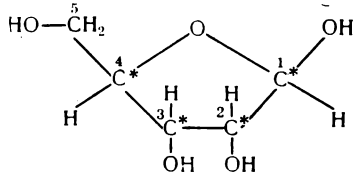


Fig. 2.3. The primary structure of sperm whale myoglobin.

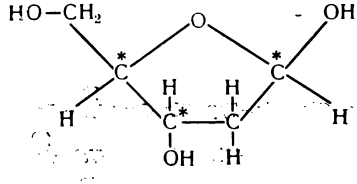
The principal chain of a nucleic acid consists of alternating phosphoric-acid units and a sugar, either the carbohydrate ribose in ribonucleic acid (abbreviated to RNA) or deoxyribose in deoxyribonucleic acid (DNA). Attached to each of the carbohydrate unit is one of the four nitrogenous bases. DNA and RNA are transcripts written in the language of the four-letter alphabet. The general scheme of the chain is as follows (P is phosphate):



The structural formulas of ribose and deoxyribose are:



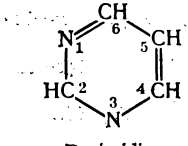
D-Ribose



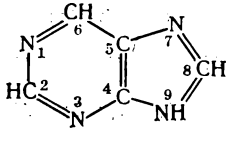
D-2-Deoxyribose

The only difference is the absence of an oxygen atom in deoxyribose at the second carbon atom. The asterisks indicate asymmetric carbon atoms (see Sec. 2.7).

The nitrogenous bases are the derivatives of pyrimidine and purine—nitrogen-containing heterocyclic compounds:



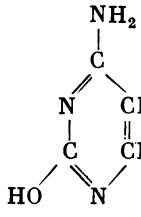
Pyrimidine

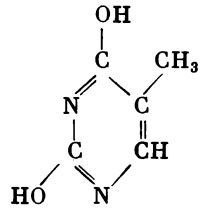


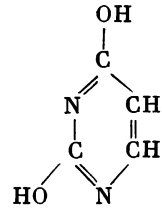
Purine

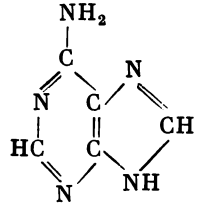
Four different bases are commonly found in each type of nucleic acid. These are adenine (A), guanine (G), cytosine (C), and thy-

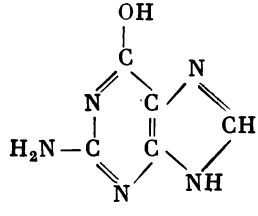
mine (T) in DNA, and adenine, guanine, cytosine, and uracil (U) in RNA. Uracil, thymine, and cytosine are the pyrimidine bases, and adenine and guanine are the purine bases.


Cytosine


Thymine


Uracil

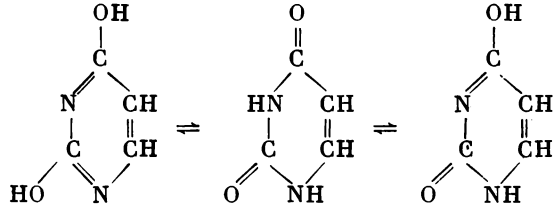

Adenine


Guanine

Apart from these commonly occurring bases, other bases have also been detected in much smaller amounts in DNA and various types of RNA, such as 5-methylcytosine, 5-hydroxymethylcytosine, hypoxanthine (guanine minus the NH_2 group, abbreviated as I), and xanthine (guanine with the NH_2 group being replaced by OH, denoted as X).

The pyrimidines and purines possess basic properties—their nitrogen atoms can accept protons, acquiring a positive charge.

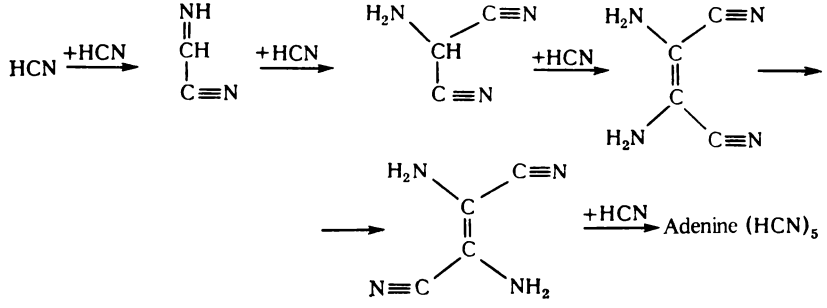
All the nitrogen bases mentioned display *tautomerism*—they can exist in several forms, called tautomers, which arise as a result of the transfer of the hydrogen atoms of the hydroxyl group, OH, to the nitrogen atoms. The tautomers of uracil are as follows:



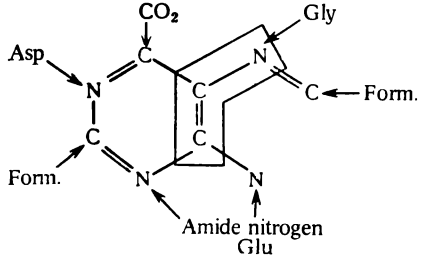
The dynamic equilibrium of tautomers, being dependent on temperature, is shifted to the keto form under ordinary conditions.

Evidently, the nitrogen bases appeared at the early stages of prebiological evolution, and the parent compound could have been hydrocyanic acid, HCN. The primary synthesis of adenine might

have occurred according to the following scheme:



The metabolic origin of the purines in living organisms is shown in the following scheme:



The abbreviation Form stands for formic acid, HCOOH. The primary formation of ribose is associated with formaldehyde, H_2CO.

Combination of a base (either a purine or pyrimidine) with a sugar (ribose or deoxyribose) gives rise to a nucleoside—a ribonucleoside (a riboside) or deoxyribonucleoside (a deoxyriboside), respectively. Cytidine, thymidine, uridine, adenosine, and guanosine are all nucleosides. The manner in which a nitrogenous base is united with a sugar is shown in Fig. 2.4. The nucleosides phosphorylated at positions 3′ and 5′ of the sugar are, respectively, called 3′-nucleoside (or 5′-nucleoside) phosphates or nucleotides (see Fig. 2.4).

The formation of a nucleic acid takes place by way of the polycondensation of nucleoside triphosphates (CTP, etc.). When a nucleotide is incorporated into the chain, a phosphate ester linkage is formed between the two nucleotides with the concomitant release of a molecule of diphosphate—pyrophosphoric acid:

$$\text{CTP} + \text{GTP} \rightarrow \text{CMP} - \text{GMP} + H_4P_2O_7$$

(here TP is triphosphate, and MP is monophosphate). Figure 2.4 shows the structure of the DNA chain. The RNA chain is built up

in a similar way, the only differences being that uracil (U) replaces thymine (T) and the OH group at C-2 of ribose is replaced by H.

The nucleotide composition of DNA obeys *Chargaff's rules*—the amount of adenine is always equal to the amount of thymine, and

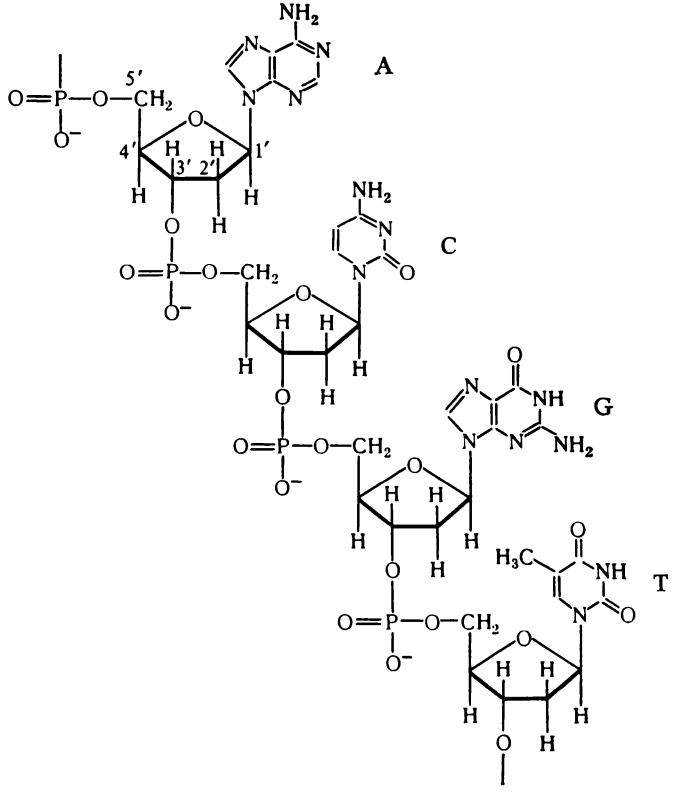


Fig. 2.4. The DNA chain.

the amount of guanine is always equal to the amount of cytosine:

$$\frac{A+G}{T+C} \approx 1$$

Thus, the total amount of the purine bases in DNA is always equal to the total amount of pyrimidine bases. At the same time,

$$\frac{T+G}{A+C} \approx 1$$

that is, the amount of 6-NH_2 groups in the bases of DNA is always equal to the number of 6-C=O groups. The ratio

$$\frac{G+C}{A+T} \approx \frac{G}{A} \approx \frac{C}{T}$$

is, however, different from unity. It varies within wide limits in bacteria (from 0.45 to 2.8) and is approximately equal to 0.45 in higher animals and plants. The species specificity of the nucleotide composition of DNA has been studied in detail by Belozersky and

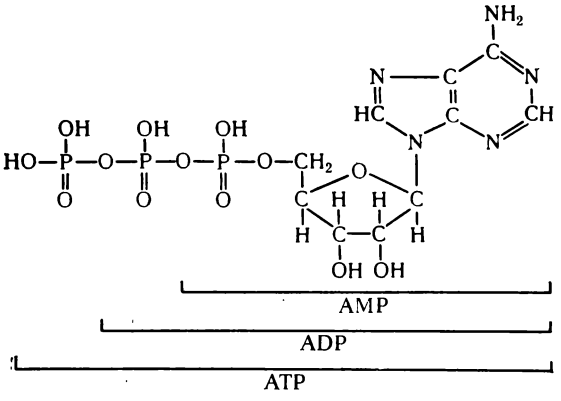


Fig. 2.5. Adenosine mono-, di- and triphosphates.

his school. The species specificity is expressed not only by the relative content of G + C but also by the content of minor bases.

Ribonucleic acid does not obey Chargaff's rules. The ratio (G + + C)/(A + U) varies within wide limits for any living organism.

Deoxyribonucleic acid (DNA), the substance of the genes, is contained in chromosomes and cell mitochondria and also in bacteriophages. The molecular weights of DNA amount to 10^9, the largest of the molecules known. Ribonucleic acid (RNA) occurs in various forms, in the nucleus, in the cytoplasm of the cell, and also in viruses and phages. Four types of RNA are known to exist: high-molecular-weight RNA, namely ribosomal RNA (rRNA) with molecular weights of the order of 10^6 and messenger RNA (mRNA) with molecular weights ranging from 30,000 upwards. Since the average molecular weight of a ribonucleotide is 224, the shortest chains of mRNA contain about 150 nucleotides. The third type of RNA, transfer RNA (tRNA), has a molecular weight of the order of 20,000 and contains

about 80 nucleotides. The fourth type is high-molecular-weight RNA associated with virus particles.

There have been devised excellent biochemical methods for the determination of the primary structure, i.e., the sequence of nucleotides in RNA and DNA. Especially thoroughly studied are the primary structures of various tRNA's; at present, the "texts" of a number of genes, i.e., segments of DNA, are already known. Moreover, some genes have been synthesized—the first synthesis was accomplished by Khorana in 1970.

2.6. Adenylates

Monomeric phosphorylated nucleosides play a very important role in metabolism and bioenergetics, in the regulation of life activity at the molecular level. This is striking evidence of the chemical unity of all forms of life (page 26) and of the diverse uses of the same substances by cells. A special place among nucleosides is occupied by adenine. Figure 2.5 shows the structure of adenosine-5′-mono-, di-, and triphosphates (AMP, ADP, ATP). Adenosine triphosphate, a so-called "energy-rich" compound, is the principal source of chemical energy in the cell. This energy is released in the course of the hydrolytic cleavage of the γ-phosphate in the reaction $ATP \rightarrow ADP + P_i$ (P_i is orthophosphoric acid, H_3PO_4, inorganic phosphate). The energy of ATP is consumed for all cell functions: for protein biosynthesis, for the active transport of substances through membranes, for the performance of mechanical and electrical work, etc. The free energy released by ATP upon elimination of the phosphate is $\Delta G^0 = 30.7$ kJ/mole. The free energy of the other nucleoside triphosphates, primarily GTP, is used less frequently. Of course, all reactions that utilize the energy of ATP, involve enzymes, ATPases, which are regulated by small ions in a number of cases. The energy released in exergonic reactions is stored by the conversion of ADP to ATP: $ADP + P_i \rightarrow ATP$; this reverse reaction takes place in the processes of glycolysis, in photosynthesis, and during respiration (oxidative phosphorylation). All these processes involve the oxidation and reduction of the important coenzymes NAD and NADP (nicotinamide adenine dinucleotide and NAD phosphate). The compounds NAD and NADP are adenylates—the derivatives of adenine. Figure 2.6 shows the interconversions of NAD and NADH—the reaction $NADH + H^+ \rightleftarrows 2H + NAD$ (the arrow indicates the hydroxyl group esterified by the phosphate to NADP and NADPH).

The metabolism of many compounds requires their conversion to the activated derivatives of coenzyme A (CoA), the most important of which is acetyl coenzyme A (acetyl-CoA) shown in Fig. 2.7. Acetyl-CoA is a carrier of acetyl groups, just as ATP is a carrier of phosphate groups. Acetyl-CoA is structurally similar to ATP.

Like the terminal phosphate bonds in ATP and ADP, the thioester linkage $-CH_2-S-CO-$ is a high-energy bond—it is an exergonic bond, i.e., a bond which when cleaved liberates a free energy. The magnitude of ΔG° for the hydrolysis of acetyl-CoA (the reaction acetyl-CoA + $H_2O \rightarrow$ acetic acid + CoASH) is -31.5 kJ/mole.

A particularly important regulatory role is played by another adenylate, *cyclic adenosine monophosphate* (cyclic AMP or c-AMP)

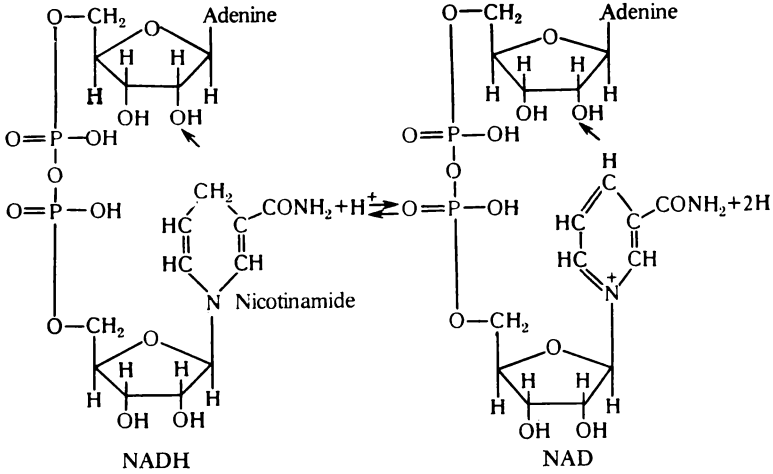


Fig. 2.6. Nicotineamide adenine dinucleotide (NAD).

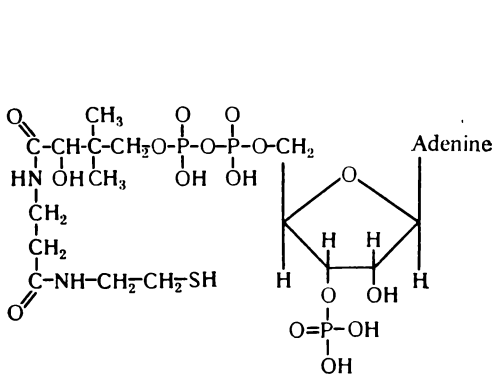


Fig. 2.7. Acetyl coenzyme A (acetyl-CoA).

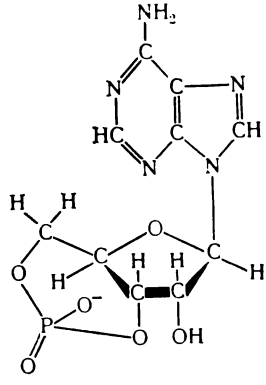


Fig. 2.8. Cyclic adenosine monophosphate.

(Fig. 2.8). Cyclic AMP controls enzymatic reactions in cells, in which sugars and fats are stored. Cyclic AMP also regulates the transcription of genes (i.e., the biosynthesis of proteins) and serves as a regulator—a universal mediator during the action of a number of hormones, in particular, adrenaline (Fig. 2.8). Cyclic AMP is present in cells in amounts three orders smaller than ATP. It is formed from ATP and, in turn, is converted into an inactive noncyclic AMP. These small quantities of c-AMP are still larger by two or three orders than the amount of the hormone and enhance more than 100-fold the external hormonal signal that acts on the cell.

2.7. The Chirality of Biological Molecules

The overwhelming majority of complex molecules containing more than three atoms have neither plane nor centre of symmetry. Such molecules are dissymmetric, or chiral. The word chiral (the Greek word *cheir* meaning "hand", pronounced kiral; compare chiropractic, chiromancy) implies the property of "handedness", i.e., the property of nonidentity of an object with its mirror image (the object and mirror-image relationship of a left and right hand). Chiral substances can exist in two forms—the left-handed and the right-handed form. These two configurations cannot be superimposed by rotations of the system as a whole in space: they are related to each other in the same way as a left hand is related to a right hand (Fig. 2.9). (This fact becomes obvious when one attempts to put a "left-handed" glove on a right hand or vice versa.)

Chirality is often encountered in the world of molecules; it is determined by a so-called asymmetric carbon atom (by convention chiral carbons are often designated with an asterisk, see p. 38). In saturated (aliphatic) organic compounds, the four valence bonds of carbon are arranged at tetrahedral angles relative to one another. If two of the four bonds join identical groups, as, for example, in the molecule CX_2YZ, the CYZ plane is then a plane of symmetry and the molecule is not chiral. A carbon atom is asymmetric (chiral) when it has four different groups attached to it (say, C*XYZV). Such a molecule has neither plane nor centre of symmetry. Thus, the property of chirality is characteristic of 19 common amino acids [all amino acids, except glycine ($H_3N^+ \cdot CH_2 \cdot COO^-$)]. Figure 2.10 shows the right-handed (D) and the left-handed (L) configuration of alanine (*mirror-image antipodes* or *enantiomers*).

When two symmetric molecules are synthesized in the laboratory, a 50 : 50 mixture of both enantiomers always results; such a mixture of enantiomers is called a *racemic mixture* or *racemate*. This follows from the second law of thermodynamics—a racemate corresponds to the maximum entropy of mixing.

A remarkable property of all forms of life is the fixation in organisms of all the most important biological molecules, beginning with the amino acids, in a single definite configuration. The amino-acid residues in proteins always have a left-handed, L-configuration (Fig. 2.10).

The two enantiomers react identically with symmetric (achiral) molecules. They differ in their interaction with polarized light. Chiral compounds, unlike their racemic mixtures, rotate the plane of polarized light in opposite directions (by equal amounts) and absorb left- and right-circularly polarized light differently (see Chapter 6).

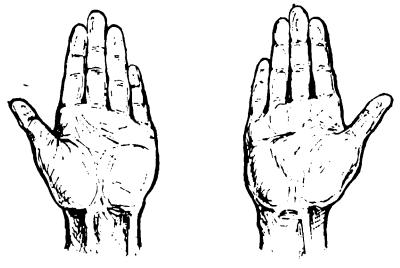


Fig. 2.9. The right and left hands.

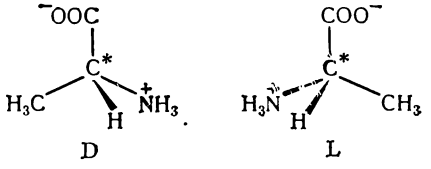


Fig. 2.10. The right-handed and left-handed configurations of alanine.

The biological L-amino acids are called so not because they rotate the plane of polarization to the left and the D-amino acids to the right. The L-amino acids may be either levorotatory or dextrorotatory. The symbols D and L show a configurational relationship to some standard substance, usually levorotatory glyceraldehyde, OCH—C*H (OH) —CH_2OH. All L-compounds can, in principle, be derived from L-glyceraldehyde by replacement of the respective atoms and groups attached to the asymmetric carbon atom (C*) without change of the overall configuration of the molecule.

The property of handedness (chirality) is exhibited by proteins, carbohydrates, nucleic acids, and also by a number of low-molecular-weight compounds in the cell. The carbohydrates in DNA and RNA are always present in the D-configuration. The nitrogenous bases have a planar structure and, consequently, display no chirality. During the metabolic processes, which involve no racemization, i.e., no interconversions of enantiomers, the cell assimilates only those compounds which correspond to the structure of its biological molecules. The organism assimilates L-amino acids but not D-amino acids. Having found its way into an environment, in which plants

and animals contain molecules with opposite configurations, an organism living on the earth will perish because of starvation. The organism responds differently to two enantiomeric forms (D- and L-antipodes). There are substances which are poisonous in one form and harmless in the other (its mirror-image form). L-Aspartic acid is tasteless, but its antipode is sweet. As far back as 1857 Pasteur observed the preferential destruction by some microbes of one of the antipodes of a substance.

The resolution of racemic mixtures with the separation of pure optical antipodes *in vitro* (the method known as asymmetric synthesis) is accomplished with the aid of chiral substances of biological origin (usually, alkaloids). A compound L′ is made to react with a racemic mixture (D, L):

$$(D, L)+L' \rightarrow DL'+LL'$$

The resulting compounds DL′ and LL′ are no longer enantiomers (mirror images): had the products been DL′ and LD′ they would have been enantiomers. So, the physico-chemical properties of DL′ and LL′ are different and they can be separated, for example, by crystallization.

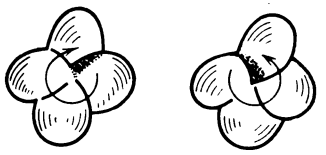


Fig. 2.11. Two forms of the shell of the rhisopodium *Neogloboquadrina pachyderma.*

To separate a racemic mixture into its two constituent enantiomers, it is necessary to make use of an asymmetric substance or living organism capable of discriminating between the levorotatory and dextrorotatory forms. Mirror-image forms were discovered by Pasteur in 1848. He examined crystals of tartaric acid and found that they were of two different forms, one the nonsuperimposable mirror image of the other. Having separated the two types of crystals with the aid of a magnifying glass and a pair of tweezers, Pasteur obtained pure optical antipodes of tartaric acid. He played the role of an asymmetric factor—the man himself is "chiral" and knows the difference between right and left. Chirality exists in living things both at the molecular level and at higher levels of organization. It is determined, in the long run, by the "chiral" control of enzymatic processes. Figure 2.11 shows two forms of the shell of the rhizopodium *Neogloboquadrina pachyderma.* Shells that are twisted in a clockwise direction are formed at a temperature lower than 7°C, those twisted in an anticlockwise direction are formed at higher temperatures.

Chirality in nature may be molecular or crystalline. In the first case it is retained upon melting or dissolution of a substance (say, a sugar) in the second—it is characteristic of the crystalline state

alone. Quartz is built up of symmetric molecules SiO_2, but it crystallizes in the levorotatory or dextrorotatory forms which possess opposite optical activities. When quartz melts, its optical activity disappears. In any deposit of quartz the number of dextrorotatory and levorotatory crystals is, on the average, identical. Accordingly, one can separate a pure optical antipode by crystallization by placing a nonracemic dextrorotatory or levorotatory seed in a racemic mixture.

The appearance and fixation of chirality in living systems are of great interest. Attempts to explain these facts by a low circular polarization of the light scattered by the earth's atmosphere or by radioactive radiation (due to the nonpreservation of parity in nuclear processes) have failed. These phenomena should be treated in the light of the general theory of prebiological evolution, which treats the emergence of order out of disorder as the creation of information (Chapter 17). The choice of an antipode means the creation of information equal to 1 bit per monomer molecule. There is weighty ground for believing that chirality first appeared as a result of fluctuations. The fluctuational departure from the equilibrium racemic distribution may increase infinitely if the system is autocatalytic, i.e., self-reproducing. We illustrate this by a model calculation. Let x_1 and x_2 be the numbers of molecules of a polymer (of the type of RNA) made up of D- and L-monomers, whose amounts are denoted as m_1 and m_2. Polymers build up their replicas out of monomers—there takes place a template self-reproduction. Moreover, polymers are capable of degradation. The kinetic equations that describe the development of the system are as follows:

$$\begin{aligned} \dot{x}_1 &= ax_1w_1 - bx_1 \\ \dot{x}_2 &= ax_2w_2 - bx_2 \end{aligned} \tag{2.22}$$

where a and b are the rate constants of polymerization and degradation, and

$$w_1 = \frac{m_1}{m_1+m_2} \quad \text{and} \quad w_2 = \frac{m_2}{m_1+m_2} \tag{2.23}$$

are the probabilities that templates 1 and 2 will encounter monomers m_1 and m_2, respectively. In the steady state $\dot{x}_1 = \dot{x}_2 = 0$, from which it follows that $w_1 = w_2 = b/a$ and, hence, $m_1 = m_2$, i.e., $w_1 = w_2 = 0.5$, $a = 2b$ and the system is racemic. The nontrivial development of the system is possible if $m_1 \neq m_2$ as a result of fluctuations. Let us assume that

$$w_1 = 0.5 + \alpha \quad \text{and} \quad w_2 = 0.5 - \alpha \tag{2.24}$$

where α is a constant. Then, the solution of the system (2.22) has the following form:

$$\begin{aligned} x_1 &= x_1(0)\ \exp\ [(a/2-b)t]\ \exp\ (a\alpha t) \\ x_2 &= x_2(0)\ \exp\ [(a/2-b)t]\ \exp\ (-a\alpha t) \end{aligned} \tag{2.25}$$

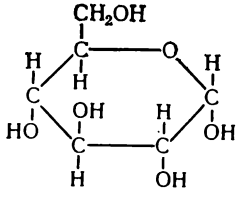
With time the population x_1 becomes dominating:

$$\frac{x_1}{x_2} = \frac{x_1(0)}{x_2(0)} \exp(2a\alpha t) \tag{2.26}$$

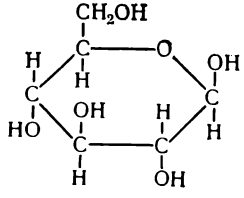
The subsequent biological evolution means further fixation of chirality, since chiral systems have advantages over racemic systems—they interact more specifically with the environment, discriminating between right and left.

2.8. Carbohydrates and Lipids

The third type of biopolymers is carbohydrates or polysaccharides built up of monosaccharide units, which have in a free state the general formula $C_6H_{12}O_6$. Monosaccharides play a most important role in the metabolism of plants and animals, particularly *glucose*:

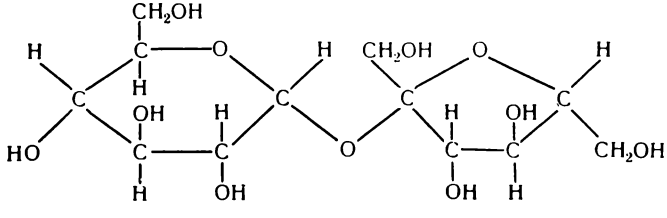


α-D-Glucose



β-D-Glucose

Each of the five carbon atoms in the ring is asymmetric. Cells also contain disaccharides, $C_{12}H_{22}O_{11}$. As an example, let us write the structural formula of sucrose (cane sugar and beet sugar):



The most important polysaccharides are the following: *starch* (two forms—*amylose* and *amylopectin*) and *cellulose* in plants, *chitin* in arthropods, and *glycogen* in animals. Cellulose and chitin are important structural and defensive materials. Starch and glycogen are substances which serve as storage forms of carbohydrates and chemical energy. Figure 2.12 shows an amylose unit. In contrast to amylose, the chains of amylopectin are branched, just as those of glycogen. Polysaccharides are not informational texts, their structure is monotonous. The size and degree of branching of these macromolecules vary within wide limits. The molecular weight of amylose that occurs in potato starch is about 10^4 (200 glucose units), and

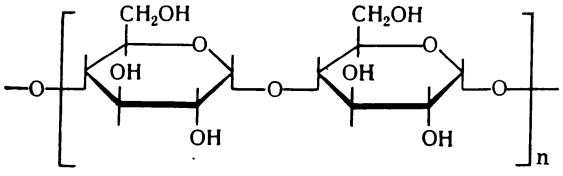


Fig. 2.12. The structure of amylose.

amylopectin present in rice starch has a molecular weight of about 5×10^5 and 80-90 branchings. The molecular weight of glycogen in muscle tissue is 10^6 and that of glycogen present in liver 5×10^6. The cotton cellulose has a molecular weight of the order of 5×10^5.

Polysaccharides play an important role in the outer membranes of some cells and are present in the cell walls of many species of bacteria. The polysaccharides contained in the membranes are found complexed with proteins and lipids.

The most important role of fatty substances, *lipids*, consists in their invariable participation in the building and functioning of biological membranes. Animal fats and vegetable oils are the triglycerides of fatty acids, i.e., esters of glycerol (1,2,3-propanetriol), such as, for example, the triglyceride of stearic acid, $H_3C(CH_2)_{16}COOH$:

$$\begin{array}{l} CH_2-O-CO-(CH_2)_{16}CH_3 \\ | \\ CH-O-CO-(CH_2)_{16}CH_3 \\ | \\ CH_2-O-CO-(CH_2)_{16}CH_3 \end{array}$$

We see that these compounds contain long "tails" of nonpolar hydrocarbon residues and strongly polar "heads" with —O—CO— groups. The functional lipids of cell membranes are more complex compounds, which contain carbohydrate, amino, and alkylamino groups. A number of important compounds belong to the class of *phospholi-*

pids. Figure 2.13 gives the structure of the phospholipid sphingomyelin. Membrane lipids and phospholipids are built up, as a rule, of a strongly polar "head" and two long nonpolar hydrocarbon "tails". The presence of unsaturated double C=C bonds in the "tails" is essential for their functioning. Vegetable oils contain such bonds but animal fats do not. The functioning of lipids in membranes will be described in Chapter 10.

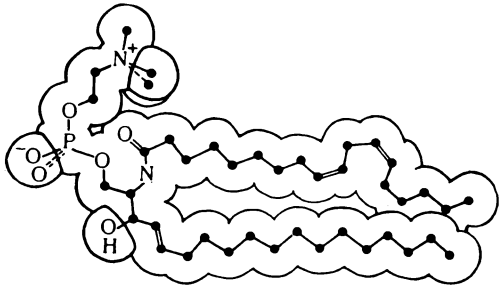


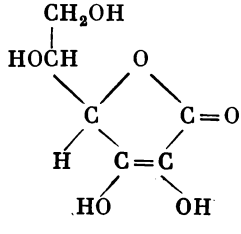
Fig. 2.13. The structure of a lipid (sphingomyelin).

2.9. Cofactors, Vitamins, and Hormones

Metabolism on the whole as well as the special properties of biopolymers in cells are regulated by specific small molecules that belong to several classes of organic compounds. The adenylates have already been described (see Sec. 2.5).

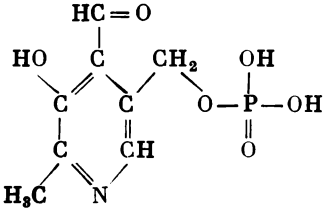
The majority of enzymes function when complexed with low-molecular-weight cofactors, called *coenzymes.* Such an enzyme is known as the *holoenzyme*, and its protein moiety is called the *apoenzyme.* There are various cofactors. To the aliphatic series belong the diphosphates of carbohydrates and their amino derivatives which take part in the transport of phosphate groups. Among aliphatic cofactors mention should be made of sulphur-containing lipoic acid and glutathione.

The necessary participant of a number of oxidation-reduction processes is ascorbic acid or vitamin C:

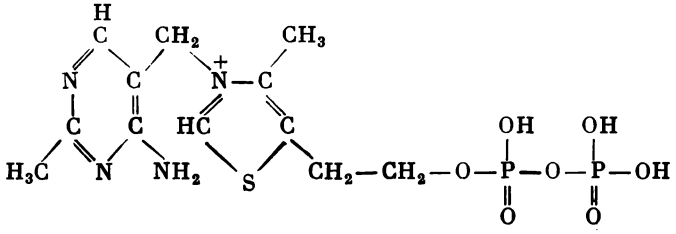


Vitamins are necessary just because they serve as cofactors or are converted to them. Since cofactors function as catalysts, they are needed in small quantities in the organism.

Unlike proteins and nucleic acids, not to mention polysaccharides, whose main chains do not contain conjugated π-bonds, the majority of the most important coenzymes are π-electronic conjugated systems containing aromatic rings or hetero rings. As we have already seen, the adenylates are such systems. The flavin coenzymes—flavin mononucleotide (FMN) and flavin adenine dinucleoside (FAD)—contain a conjugated three-ring heterocyclic compound—riboflavin (vitamin B_2). These coenzymes are yellow-orange in colour since riboflavin absorbs light in the visible region of the spectrum. Adenine enters into the composition of coenzymes associated with cobalamin or vitamin B_{12}. Pyridoxal phosphate, a derivative of pyridine, is the coenzyme of transaminases, which catalyze the conversions of amino acids. The structural formula of pyridoxal phosphate is



Thiamine phosphate too contains a π-electronic system:

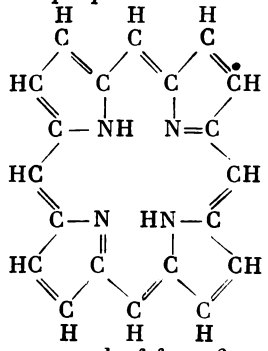


Thiamine is vitamin B_1.

A long linear conjugated system of π-bonds is contained in β-*carotene*, a yellow pigment in carrots, which is a precursor of vitamin A. Its modification, *retinal*, plays a key role in the chemistry of vision (Chapter 14).

Of prime importance for a number of life processes, starting with photosynthesis, are the π-electronic conjugated systems of *porphyrin*

compounds—derivatives of *porphin*:



The porphyrin ring is composed of four five-membered pyrrole hetero rings. Like other conjugated systems, the porphyrin ring is flat. Porphyrins easily form chelates with many metal ions, the metal being coordinately linked to the four nitrogen atoms in the centre of the planar system. The cobamide enzymes mentioned above contain a porphyrin ring system with Co in the centre. Chlorophylls, which play an essential role in the process of photosynthesis, contain a porphyrin with the central Mg atom (see Sec. 14.1). The prosthetic group heme, an iron porphyrin complex, is present in proteins that participate in the respiratory process—cytochromes (see Sec. 13.6), in myoglobin and hemoglobin (see Sec. 6.7), in the legoglobin of nodule bacteria which participates in the fixation of atmospheric nitrogen, and in a number of oxidation-reduction (redox) enzymes. The blood of the crustaceans contains porphyrins with the central Cu atom, and half of porphyrin, a dipyrrole complex of V is present in the blood of the tunicates (*Ascidia*).

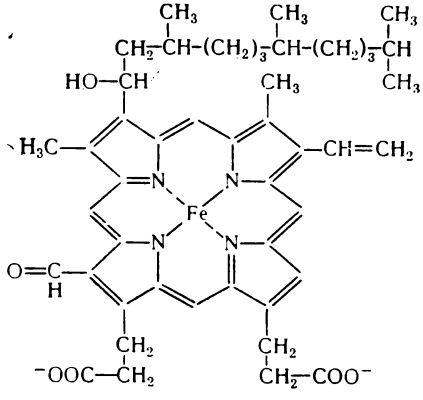


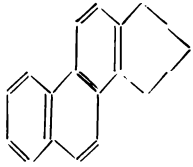
Fig. 2.14. The heme (protoheme IX).

Figure 2.14 shows the structure of heme, the prosthetic group of myoglobin and hemoglobin.

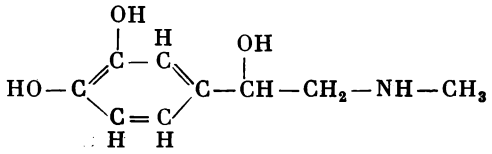
The role of metals, in particular the complex-forming transition metals, is very essential in biology. About one-third of all the enzymes known to science contain metal ions as cofactors. As has been said above, the photosynthesis involves Mg and also Fe (cytochromes, ferredoxin); the storage and transport of oxygen and

oxidation-reduction electron-transfer processes are also associated with heme. The ions of alkali and alkaline-earth metals take part in protein biosynthesis; Zn^{2+} ions participate in the synthesis of nucleic acids (in *polymerases*) and in other enzymatic processes. The transmission of the nervous impulse is directly controlled by the transport of K^+ and Na^+ ions through membranes; and the mechanochemical processes (say, muscular contraction) require the participation of Ca^{++} and Mg^{++}. Nitrogenase, an enzyme responsible for the fixation of nitrogen, contains Mo atoms. The study of the substances mentioned above and a number of other biological substances and also of biological processes relies on modern inorganic chemistry. The main objects studied in this field are complex compounds of metals. *Bioinorganic chemistry*, which has been developed in the last several decades, provides much information for biology and biophysics.

In complex, multicellular organisms, the role of "signalling" and regulatory substances is played by *hormones*. There are two major groups of hormones in animal organisms—proteins, polypeptides and their derivatives, and steroids. The first group includes *thyroglobulin*, the protein of the thyroid gland, which contains iodated thyroxin; *insulin*, which regulates the sugar content in the blood; *oxytocin*, which causes contraction of the uterus; *vasopressin*, which regulates the blood pressure, etc. The hormones are synthesized in the glands of internal secretion (the endocrine glands) and effect regulation at the organism level. Steroids are compounds that contain the carbon skeleton of cyclopentanophenanthrene.



The most important steroid hormones are the sex hormones: estrone, progesterone, testosterone, androsterone, and cortisone, a hormone of the adrenal cortex. Apart from these two classes, there are other low-molecular-weight hormones. For example, adrenalin (epinephrine), a hormone of the adrenal medulla, which is involved in the function of the nervous system and causes such effects as an increase in blood pressure and an acceleration of the heart beat:



The hormonal activity is controlled by the chemical functionality of a few atomic groupings. The chemical differences between testosterone and cortisone are insignificant, but their physiological functions are quite different. The structure and properties of hormones demonstrate the biological importance* of individual molecular structures.

An important role is played by the oligopeptides of the mammalian brain. These substances, which contain a small number of amino acids (for example, undecapeptide, H-Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2, which acts like morphine), have various functions—to this class belong analgetics and sleep and memory stimulants.

Substances of the polypeptide type include a number of neurotoxins present in the poisons of snakes, scorpions, etc. These substances block the synaptic transmission of the nervous impulse (see Sec. 11.5).

In concluding this section, we give a table which shows the variety of substances that function in the bacterial cell of *Escherichia coli*.

Table 2.3. The Approximate Chemical Composition of the Cells of *Escherichia coli*

Substance	Average molecular weight	Number of molecules in cell	Number of different molecules
H_2O	18	10×10^{10}	1
Inorganic ions (Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, PO_4^{3-}, CO_3^{2-}, etc.)	40	2.5×10^8	20
Carbohydrates and their precursors	150	2×10^6	200
Amino acids and their precursors	120	3×10^7	100
Nucleotides and their precursors	300	1.2×10^7	200
Lipids and their precursors	750	2.5×10^7	50
Other small molecules	150	1.5×10^7	200
Proteins	4×10^4	10^6	$2\text{-}3\times10^3$
DNA	2.5×10^9	4	1
RNA:			
rRNA 16S*	5×10^5	3×10^4	1 (?)
rRNA 23S	10^6	3×10^4	1 (?)
tRNA	2.5×10^4	4×10^4	40
mRNA	10^6	10^3	10^3

* S is the symbol for Svedberg units (see page 87).

2.10. The Main Biochemical Processes

All the compounds briefly described above take part in the life processes of various organisms, which have developed in the course of biological evolution. Biosynthesis, metabolism, and bioenergetics are accomplished by large and small biological molecules.

As has already been said, nucleic acids are responsible for protein biosynthesis. This is their sole "legislative" function. In their turn, the "executives"—proteins—are the requisite participants in all biochemical processes and serve as catalysts, i.e., function as enzymes.

For biosynthesis and metabolism to be accomplished, the energy is stored in cells in a chemical form, primarily in the exergonic third and second phosphate bonds of ATP. Accordingly, the metabolic bioenergetic processes result in "charging the storage battery"—the synthesis of ATP from ADP and inorganic phosphate. This takes place in respiration and fermentation processes. The currently living organisms "remember" about the evolution that began about 3.5×10^9 years ago. There is convincing evidence that life on the earth originated in the absence of free oxygen (see Sec. 17.2). The metabolic processes that occur in the presence of oxygen (primarily, the oxidative phosphorylation during respiration) are relatively few in number; evolutionally, such metabolic processes occurred at a later time than did anaerobic processes. No complete burnout (oxidation) of organic molecules of nutrient substances is possible in the absence of oxygen. Nevertheless, as shown by the properties of the currently existing anaerobic cells, the energy required for life-sustaining processes is produced in them too in the course of oxidation-reduction processes. In aerobic systems, the final acceptor (i.e., the oxidizing agent) of hydrogen is O_2, and in anaerobic systems the role of the final hydrogen acceptor is played by other substances. Oxidation in the absence of O_2 proceeds by two ways of fermentation—glycolysis and alcoholic fermentation. *Glycolysis* involves the multistep breakdown of hexoses (say, glucose) down to two molecules of pyruvic acid, each containing 3 carbon atoms. On their way 2 molecules of NAD are reduced to NADH and 2 molecules of ADP are phosphorylated—there are produced 2 molecules of ATP. The reverse reaction NADH → NAD + H leads to the formation of lactic acid from pyruvic acid. Upon alcoholic fermentation the enzymes in yeast convert the pyruvic acid to ethyl alcohol and carbon dioxide. The overall reaction is as follows:

$$C_6H_{12}O_6 + 4NAD + 2ADP + 2P \rightarrow \underset{\text{Pyruvic acid}}{2H_3C \cdot CO \cdot COOH} + 4NADH + 2ATP$$

$$2H_3C \cdot CO \cdot COOH + 4NADH \rightarrow \underset{\text{Lactic acid}}{2H_3C \cdot CH(OH) \cdot COOH} + 4NAD$$

$$2H_3C \cdot CO \cdot COOH + 4NADH \rightarrow 2C_2H_5OH + 2CO_2 + 4NAD$$

The secondary step of the glycolytic and alcoholic fermentation is the so-called pentose-phosphate pathway (also known as the phosphogluconate pathway or the hexose monophosphate shunt) which

involves the decomposition of water and also the evolution of carbon dioxide.

The presence in the environment of CO_2, which is formed in the course of anaerobic oxidation, makes possible *photosynthesis*. The bacterial photosynthesis and the photosynthesis of green plants began about 3×10^9 years ago. Photosynthesis consists of the absorption of light and the transformation of its energy to the chemical energy of biochemical molecules. For the photosynthesis to occur light-absorbing compounds were required, in particular those containing porphyrin rings—chlorophyll and cytochromes. As a result of the absorption of light quanta in chlorophyll, the electrons of the system move to higher energy levels. Then the electron-transport chain is operative, in which the principal participants are oxidation-reduction enzymes—cytochromes. The energy stored originally in chlorophyll is released in a biologically useful form—in ATP and NADP. There takes place photophosphorylation.

Evidently, at a later stage of evolution carbon dioxide was reduced photosynthetically to carbohydrates. The hydrogen for procaryotes is supplied by various molecules (bacterial photosynthesis) and for green plants by H_2O. The oxygen is released into the surroundings. Its accumulation in the atmosphere is the result of photosynthesis. The overall reaction is

$$6CO_2 + 12H_2O \rightarrow C_6H_{12}O_6 + 6H_2O + 6O_2$$

About $1.5\text{-}2 \times 10^9$ years ago the partial pressure of oxygen in the atmosphere was 0.02-0.20 percent of the present-day level. This led to aerobic metabolism, respiration. The cellular respiration involves a number of interconnected processes leading to the synthesis of biological molecules required for life, and the "charging" of ATP (oxidative phosphorylation). The molecules of nutrients are "burned", oxidized to CO_2 and H_2O, the oxygen molecule serving as the final hydrogen acceptor. Roughly speaking, the liberation of chemical energy from food occurs in three steps. The first step is the breakdown of macromolecules and fatty substances. Proteins give rise to amino acids, carbohydrates (starch and glycogen) produce hexoses, and fats form glycerol and fatty acids. At the second step these substances give rise to acetyl-coenzyme A (page 44), α-ketoglutarate, and oxaloacetate. The third stage involves the *Krebs cycle* or the *citric acid cycle*. The Krebs cycle begins with the reaction between oxaloacetic acid and acetyl-CoA and the oxaloacetic acid is then regenerated during the second revolution of the cycle. Some of the reactions involved in the cycle, which perform a catalytic function, are accompanied by the transfer of hydrogen atoms to NAD or flavoprotein. For every 2 electrons transported with NADH to O_2 there are synthesized 3 molecules of ATP (*oxidative phosphorylation*). This electron transport is accomplished in the respiratory chain. Upon

transition from the first to the second phase carbon dioxide is released, and the transport of electrons from NADH to molecular oxygen is followed by the formation of water. The general scheme of aerobic respiration is

$$C_6H_{12}O_6 + 6H_2O + 6O_2 + 36ADP + 36P \rightarrow 6CO_2 + 12H_2O + 36ATP$$

Respiration is much more efficient than glycolysis. Upon glycolysis the free energy change per mole of glucose is 197 kJ, whereas upon respiration it is 2880 kJ.

The complex processes of metabolism, accumulation, and consumption of energy are spatially localized in cells. Respiration occurs in mitochondrial membranes and photosynthesis takes place in the membranes of chloroplasts. Biochemical processes are evolutionally adapted. For instance, the main source of metabolic energy for animals living in deserts and for birds is glycogen rather than fats. In deserts not only the maximum release of energy but also the maximum formation of water must be provided: the oxidation of fats produces twice as much water as the oxidation of glycogen. That is why birds have a smaller mass of fat. The mass of glycogen and of the associated water is 8 times greater than the mass of the fat that liberates on oxidation the same amount of energy.

A detailed consideration and schemes of all the metabolic pathways are given in books on biochemistry. We shall return to respiration and photosynthesis in Chapter 13 which is devoted to bioenergetics and in Chapter 14 dealing with photobiological processes.

2.11. Strong and Weak Interactions

The interactions of atoms in biological molecules as well as in molecules of synthetic organic compounds are, first of all, chemical covalent bonds, which we shall call strong interactions. The energy required for the rupture of a C—C bond is equal to 348.6 kJ/mole, the C—N bond energy is 336 kJ/mole, etc. Strong interactions determine the chain structure of biopolymers, the binding of monomers—amino acid residues, nucleotides, hexoses. Strong bonds are formed by the outer electrons of atoms; the theory of covalent bonds can be based only on quantum mechanics. The corresponding branch of physics or theoretical chemistry is known as *quantum chemistry.*

In the course of biochemical reactions catalyzed by enzymes there takes place a rearrangement of chemical bonds and of electron shells.

Biological molecules would not be able to function and all known forms of life would not be able to exist at all without nonvalent, nonchemical, weak forces operating inside biological molecules and between them in addition to strong interactions. Cells and their organelles are heterogeneous systems, whose existence and functioning are determined by the intermolecular interactions of nonva-

lent character. Proteins, which perform nearly all molecular functions in cells, interact with lipids and carbohydrates, with nucleic acids, and with small molecules. These interactions are predominantly weak, since strong interactions would have built up too rigid and stable structures devoid of molecular mobility required for biological molecules to accomplish their diverse tasks, including the delicate control of chemical reactions, compartmentation, and the setting-up of concentration gradients. Let us describe the types of weak interactions operative in biological systems.

1. *Ionic bonds.* The forces of ionic interactions described by Coulomb's law. The energy of interaction between two ions with charges e_1 and e_2 is given by

$$U_{\text{ion}} = \frac{e_1 e_2}{\varepsilon r} \tag{2.27}$$

where r is the distance between two ions; ε is the dielectric constant of the medium. Like charges repel each other, $U_{\text{ion}} > 0$, and unlike charges attract each other, $U_{\text{ion}} < 0$. Ionic bonds are formed, in particular, between ionogenic groups in proteins (e.g., Glu^- and Lys^+), between such groups and small counterions, between phosphate groups in nucleic acids and cations, etc.

2. *Ion-dipole interactions* between ions and polar groups of molecules. They are determined by the ionic charge and the dipole moment of the atomic group that interacts with it.

3. *Orientation forces.* The electrostatic interactions between dipoles. The dipole moment p of a small molecule or atomic group is equal, in order of magnitude, to the product of the electronic charge (4.8×10^{-10} esu) by the length of chemical bond (about 10^{-8} cm). The traditional unit for dipole moment is the debye (D); $1\text{D} = 10^{-18}$ esu. Dipoles tend to line up in an antiparallel fashion or one after another. The energy of orientational interaction between two dipoles is inversely proportional to the cube of their separation:

$$U_{\text{or}} = \frac{1}{r^3}\left\{\boldsymbol{p}_1\boldsymbol{p}_2 - \frac{3(\boldsymbol{p}_1\boldsymbol{r})(\boldsymbol{p}_2\boldsymbol{r})}{r^2}\right\} \tag{2.28}$$

If two dipoles are lined up in a "tail-to-tail" fashion, i.e., all the three vectors, $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, and $\boldsymbol{r}$ are collinear, then

$$U_{\text{or}} = -\frac{2p_1p_2}{r^3} \tag{2.29}$$

If the dipole molecules are in a state of thermal motion in a gas or in a liquid, then expression (2.28) should be averaged over all the mutual orientations of dipoles with account taken of the Boltzmann factor $\exp(-U/\text{k}T)$. At a sufficiently high temperature, when $U \ll \text{k}T$, the averaging gives

$$U_{\text{or}} = -\frac{2p_1^2p_2^2}{3\text{k}Tr^6} \tag{2.30}$$

4. *Inductive forces.* The permanent dipole of a molecule or atomic group induces in another molecule or atomic group a dipole moment with which it interacts. The dipole moment induced by an electric field of strength $\boldsymbol{E}$ is given by

$$\boldsymbol{p} = a\boldsymbol{E} \tag{2.31}$$

where a is the polarizability which characterizes the ability of the electron shell to be displaced under the influence of the electric field. The polarizability has the dimension of volume; the order of magnitude of the polarizability of atoms and small molecules is the same as that of their volumes, i.e., 10^{-24} cm^3. In our case, the strength of the electric field for a dipole in a molecule separated from it by a distance r is equal to $E = 2p/r^3$ (if $\boldsymbol{p}$ and $\boldsymbol{r}$ are parallel), and the energy of induced interaction is given by

$$U_{\text{ind}} = -\frac{1}{2} aE^2 = -\frac{2ap^2}{r^6} \tag{2.32}$$

Like U_{or}, the energy of induced interaction is inversely proportional to the sixth power of the distance and is independent of the temperature.

Formulas (2.28) through (2.32) are valid only for point dipoles, i.e., on the condition that $p \ll er$, where e is the charge of the electron. If this condition is not fulfilled, then it is necessary to make calculations of interactions of the point charges (monopoles).

The interactions mentioned above are all electrostatic; the energy of attraction or repulsion is calculated on the basis of classical electrostatics. Quantum effects are practically insignificant.

5. *Dispersion forces* (or *London forces*). The interactions of the valence-saturated electron shells of atoms and molecules. These are forces which are independent of the presence of charges, dipole moments, quadrupole moments, etc. Such are, for example, the forces operating between the molecules N_2, O_2, CO_2 and noble gases taken separately or in their mixtures (air). These forces determine the nonideal behaviour of dipoleless gases, their liquefaction. The dispersion forces are responsible for the existence of molecular crystals, in particular, crystals of hydrocarbons. The dispersion forces are quantum mechanical in origin; their existence cannot be explained within the framework of classical physics.

An elementary calculation of the energy of dispersion interaction is based on the representation of electrons as harmonic oscillators. Two identical oscillators undergo an instantaneous dipole-dipole interaction, as a result of which, instead of the original vibration with frequency ω_0, there arise two normal vibrations with frequencies that differ from ω_0 the more strongly, the stronger is the interaction. Accordingly, the zero-point energy of vibration is also changed. In

the absence of an interaction of this kind it is equal to

$$\mathscr{E}_0^0 = 6\frac{\hbar\omega_0}{2} \tag{2.33}$$

where $\hbar$ is Planck's constant. The presence of the factor 6 is accounted for by the fact that each of the electron-oscillators is assumed to have three degrees of freedom.

Upon interaction the zero-point energy of vibrations is

$$\mathscr{E}_0 = \frac{1}{2}\hbar\omega_0\left\{2\left[\left(1+\frac{e^2}{fr^3}\right)^{1/2}+\left(1-\frac{e^2}{fr^3}\right)^{1/2}\right]+\right.$$
$$\left.+\left(1+\frac{2e^2}{fr^3}\right)^{1/2}+\left(1-\frac{2e^2}{fr^3}\right)^{1/2}\right\} \tag{2.34}$$

where e is the charge on the electron; f is the coefficient of elasticity of the oscillator, and r is the separation of the oscillators. If $e^2/fr^3 \ll \ll 1$, then

$$\mathscr{E}_0 \approx 3\hbar\omega_0 - \frac{3}{4}\hbar\omega_0\frac{e^4}{f^2r^6} \tag{2.35}$$

and, hence, the interaction energy is given by

$$U_{\text{dis}} = \mathscr{E}_0 - \mathscr{E}_0^0 = -\frac{3}{4}\hbar\omega_0\frac{e^4}{f^2r^6} \tag{2.36}$$

Let us place an electron-oscillator in an electric field E. The force that acts on the electron, eE, is balanced by an elastic force fr. Hence, the induced dipole moment is equal to

$$p = er = \frac{e^2}{f}E = aE \tag{2.37}$$

where a is the polarizability of the oscillator. It is defined as

$$a = \frac{e^2}{f} = \frac{e^2}{m\omega_0^2} \tag{2.38}$$

where m is the mass of the oscillator. Substituting a into expression (2.36), we obtain:

$$U_{\text{dis}} = -\frac{3}{4}\hbar\omega_0\frac{a^2}{r^6} \tag{2.39}$$

The energy of dispersion interaction is also inversely proportional to the sixth power of the separation between the interacting systems. The presence of the factor $\hbar$ in expression (2.39) is an indication of the quantum-mechanical nature of dispersion forces, which were first discovered by Fritz London (1930).

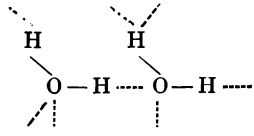
The dispersion forces are of universal value for the intra- and intermolecular interactions between atomic groups and molecules with saturated valence bonds.

Forces 3, 4, and 5 are usually termed van der Waals forces, since overall they determine the intermolecular attraction which is re-

sponsible for the correction for the pressure in the van der Waals equation. The correction for the volume is determined by the intermolecular repulsion at small distances, this repulsion being also of quantum-mechanical nature.

The order of magnitude of the energy of van der Waals interactions corresponds to the heat of vaporization of a liquid. It is of the order of 10 kJ/mole.

6. *Hydrogen bonds.* The specific donor-acceptor interactions that are induced by the hydrogen atom in groups O—H, N—H, F—H and Cl—H, and sometimes in S—H groups. The hydrogen atom binds these groups with valence-saturated oxygen, nitrogen, and fluorine atoms, say, in water:



Hydrogen bonds determine the structure and properties of water; they play an important role in the formation of the structure of biopolymers and in their interactions with small molecules. The hydrogen bond energies are of the order of 4-29 kJ/mole. They will be discussed in greater detail in Sec. 4.3.

7. *Hydrophobic interactions.* Biopolymers—proteins and nucleic acids—function mainly in aqueous medium. The hydrophobic forces are the forces of specific repulsion between nonpolar atomic groups and water molecules. They constitute the entropy effect determined by the special features of the structure of water as a condensed system. Hydrophobic interactions play an important role in the formation of the structure of proteins, supermolecular systems (membranes), etc.

Hydrophobic interactions will be considered in more detail in Sec. 4.5.

Weak interactions are responsible for molecular flexibility, i.e., the conformational properties of biopolymers and small molecules (see Chapter 3). They determine molecular recognition realized in enzyme catalysis (Chapter 6), in biosynthesis (Chapter 8), and in a number of biological processes that take place at the molecular level.

Chapter 3 Physics of Macromolecules

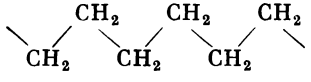
3.1. Macromolecules and Rubber Elasticity

Biological macromolecules—proteins and nucleic acids—are very complicated. Their properties in living systems are governed by the totality of structural features, in particular, by the fact that they are *informational* macromolecules and serve as "texts". It is important to establish how the behaviour of biopolymers is associated with the very fact of their chain structure, irrespective of the particular atomic groups that enter into the composition of the macromolecule. Simple noninformational chains of synthetic polymers serve as models for the investigation of this problem.

As has already been said, the physics of macromolecules is one of the foundations of molecular biophysics.

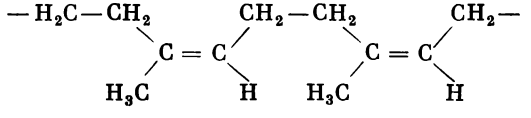
Polymers are synthesized by chemical methods or are produced from plants (caoutchouc or natural rubber, cellulose) mainly because of their valuable physical properties. In engineering, polymers are used as plastics, insulators, fibres, and rubberlike materials—natural and synthetic rubbers.

The simplest synthetic polymer is polyethylene



which is produced by the *polymerization* of ethylene, $H_2C=CH_2$: this process involves the rupture of double bonds.

The analogue of natural rubber, synthetic *cis*-polyisoprene,



is prepared by the polymerization of isoprene, $H_2C=C(CH_3)-CH=CH_2$. A number of polymers, say nylon and similar polymers, are produced by the process of *polycondensation*, which takes

place through the elimination of some low-molecular compound. As we have seen, the polypeptide chain is produced by the polycondensation of amino acids, which proceeds with the elimination of water, and a polynucleotide chain is obtained by the polycondensation of nucleoside triphosphates with the elimination of pyrophosphate.

A specific property of polymers which is the most important to biophysics is *rubber elasticity* defined as the ability of a block rubberlike polymer to undergo large elastic deformations (amounting to hundreds of percent) at a low elastic modulus. Like other elastic bodies, natural rubber obeys Hooke's law on small deformations—the stress is proportional to the strain:

$$\sigma = \varepsilon \frac{L - L_0}{L_0} \tag{3.1}$$

where σ is the stress; L and L_0 are the lengths of the stretched and undeformed samples; ε is the modulus of elasticity. For steel $\varepsilon \approx$ $\approx$ 200,000 MPa, and for rubbers $\varepsilon \approx$ 0.2-8 MPa (depending on the degree of vulcanization of rubber). In this sense natural rubber resembles an ideal gas. An ideal gas obeys the Clapeyron equation:

$$pV = RT \tag{3.2}$$

Let us compress the gas contained in a cylinder with a piston at constant temperature, increasing the pressure by an amount dp. The volume will decrease by an amount dV. From Eq. (3.2) it follows that

$$dp = -p \frac{dV}{V} = p \frac{L_0 - L}{L_0} \tag{3.3}$$

where L_0 and L are the initial and final positions of the piston, respectively. Equation (3.3) is analogous to Eq. (3.1), the only difference being that the role of the elastic modulus ε is played by the pressure p. The atmospheric pressure corresponds to ε = 100 kPa, this being of the same order of magnitude as the elastic modulus of natural rubber. The ideal gas is heated on adiabatic compression. In an analogous way, rubber becomes heated on adiabatic extension. This means that in either case the entropy decreases on deformation. The work involved in the extension of rubber by a length dL by applying a force f is given by the expression

$$f\,dL = dF = dE - T\,dS \tag{3.4}$$

where F is the free energy; E is the internal energy; S is the entropy. Rubber is practically incompressible. The elastic force on isothermal extension of rubber is equal to

$$f = \left(\frac{\partial F}{\partial L}\right)_T = \left(\frac{\partial E}{\partial L}\right)_T - T\left(\frac{\partial S}{\partial L}\right)_T \tag{3.5}$$

Experiment shows that for rubber the force f is proportional to T: the straight line representing the dependence $f(T)$ passes near the coordinate origin. In other words,

$$\left(\frac{\partial E}{\partial L_i}\right)_T \approx 0 \quad . \tag{3.6}$$

Just as the internal energy of the ideal gas is independent of the volume

$$\left(\frac{\partial E}{\partial V}\right)_T = 0 \tag{3.7}$$

so the internal energy of rubber does not depend on the length. The elastic force is determined in either case by the change in entropy rather than in internal energy. For rubber

$$f \approx -T\left(\frac{\partial S}{\partial L}\right)_T \tag{3.8}$$

In this lies the fundamental difference between the rubber elasticity of a polymer and the elasticity of a solid body (say, a steel spring), the latter being determined by the change in internal energy.

But why does natural rubber resemble the ideal gas? The entropy nature of the elasticity of an ideal gas implies that when the volume of the gas decreases, the number of impacts of molecules against the walls increases—the elastic force is determined by the thermal motion of molecules. The compression of the gas reduces its entropy, since the gas passes from a more probable rarefied state to a less probable compressed state. Therefore, the modulus of elasticity of an ideal gas is proportional to absolute temperature:

$$p = \frac{R}{V}T$$

The fact that the modulus of elasticity of natural rubber is proportional to absolute temperature, which follows from Eq. (3.8), is also an indication of the entropy nature of rubber elasticity, indicating that natural rubber is composed of a large number of independent elements subject to thermal motion. The transition from the more probable arrangement of these elements to the less probable arrangement occurs on extension of rubber. The analogy between the properties of rubber and those of an ideal gas may consist only in what has been said. But what are these elements? What plays the role of gaseous molecules in natural rubber?

The difference between macromolecules and small molecules consists primarily in that macromolecules are built up of a large number of identical links united into a linear chain. As a rule, macromolecules contain single σ-bonds, such as C—C, C—N, C—O. There are possible rotations of atomic groups about these bonds. The rotations

about single bonds give rise to various conformations of the chain. A macromolecule possesses a conformational lability, various degrees of flexibility. The role of independently moving elements is played by chain segments, which perform independent rotations. As we shall see, the conformational properties of biological molecules are very important. A study of these properties must be started with natural rubber. Following this path, we shall, in the final analysis, come to the understanding of the nature of enzymic activity.

Landau said that the objective of physics is to establish new relationships between seemingly unrelated, remote phenomena. We have already established what relates natural rubber to an ideal gas. Further on we shall see what relates an enzyme to natural rubber.

3.2. Internal Rotation and Rotational Isomerism

Classical organic chemistry considered the rotations of atomic groups about single bonds to be free. Any *conformations*, say, of ethane, $H_3C—CH_3$, that arise as a result of internal rotations have the same energy; no energy is required for the angle of rotation to be changed. Some of the conformations of ethane are shown in Fig. 3.1.

Studies of the thermodynamic properties of ethane and other compounds with single bonds and also structural investigations carried out by spectroscopical methods, NMR, etc., have, however, demonstrated that internal rotation is not free in nearly all cases. The ethane molecule has a lowest energy in the staggered or *trans*-conformation (Fig. 3.1*a*) and a highest energy in the fully eclipsed conformation (*cis*-conformation). For the C—C bond to be rotated by 120°, i.e., for the transition of one *trans*-conformation to an identical one, an energy barrier equal to 12,200 J/mole must be overcome. For ethane and also for other molecules with an axis of symmetry of C_3, the dependence of the potential energy of the molecule, U, on the angle of rotation φ can be expressed approximately by the formula

$$U = \frac{1}{2} U_0 (1 - \cos 3\varphi) \tag{3.9}$$

where U_0 is the height of the potential barrier (Fig. 3.2). The value of U_0 increases when the hydrogen atoms in ethane are replaced by more bulky atoms and groups and decreases when the axis of rotation is elongated (cf. $H_3C—CH_3$, $H_3C—SiH_3$, $H_3Si—SiH_3$). Some experimental values of U_0 are listed in Table 3.1.

The potential energy of internal rotation is determined by the weak interactions of nonbonded atoms and groups (in ethane, this is the interaction of the hydrogen atoms of one CH_3 group with those of the other CH_3 group). A rigorous quantum-mechanical calculation

Table 3.1. The Heights of Potential Barriers to Internal Rotation

Molecule	U_0, kJ/mole	U_0, kcal/mole	Molecule	U_0, kJ/mole	U_0, kcal/mole
H_3C-CH_3	12.2	2.9	H_3C-OH	4.6	1.1
$H_3C-CH_2CH_3$	14.3	3.4	H_3C-OCH_3	11.3	2.7
$H_3C-CH(CH_3)_2$	16.4	3.9	H_3C-SH	5.5	1.3
$H_3C-C(CH_3)_3$	18.5	4.4	$H_3C-CH=CH_2$	8.4	2.0
H_3C-CH_2F	13.9	3.3	H_3C-SiH_3	7.1	1.7
H_3C-CH_2Cl	15.5	3.7	$H_3Si-SiH_3$	4.2	1.0
H_3C-CH_2Br	15.1	3.6	$H_3C-C\equiv C-CH_3$	0	0

of U_0 is a formidable task, since U_0 is much smaller than the total energy of chemical bonds in the molecule and is defined as the minor difference between two large quantities—the total energies of the molecule in the *cis*- and *trans*-conformations. An energy barrier appears because of the repulsion of nonbonded atoms and of the interaction of bonds adjacent to the axis of rotation (the bond orientation

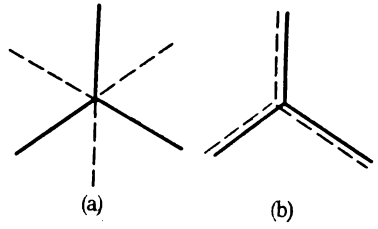


Fig. 3.1. The C—H bonds of ethane in (*a*) *trans*- and (*b*) *cis*-conformations. The projection onto the plane perpendicular to the C—C bond.

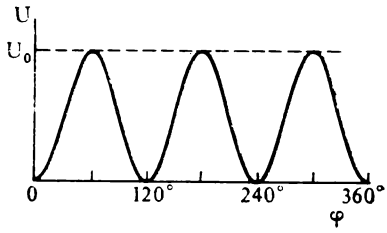


Fig. 3.2. The potential energy of internal rotation in ethane versus the angle of rotation.

effects). These two factors make the *trans*-conformation more stable (the principle of staggered bonds). The interatomic repulsion can be estimated if we know the characteristic dependences of the energy of intermolecular interaction on the distance between molecules for model substances (see page 68). It is more difficult to evaluate the bond orientation effect. A crude estimate can be obtained if we assume that in ethane, because of the small van der Waals radius of the hydrogen atom, the function $U(\varphi)$ is entirely determined by the orientational effect and that the same effect is operative in ethane

derivatives. Then we obtain for the derivatives of ethane:

$$U(\varphi) = \frac{1}{2} U_{\text{ethane}} (1 - \cos 3\varphi) + \sum_{i,j} U(r_{ij}) \qquad (3.10)$$

The quantity $U(r_{ij})$ represents the energy of van der Waals interactions between nonbonded atoms i and j separated by a distance r_{ij}. It is clear that r_{ij} depends on φ.

If the bonds are nonpolar, then $U(r_{ij})$ is composed of the energy of dispersion forces (page 60) and the energy of repulsion.

If the interaction involves polar molecules or groups possessing dipole moments, then, along with dispersion forces, there operate electrostatic forces—orientational and inductive forces (page 59).

All the three types of forces depend on the distance in the same manner—the corresponding energies are inversely proportional to r^6. The energies of van der Waals repulsion depend on the distance even more strongly, since the repulsion is substantially stronger at small distances. It may be assumed in an approximate way that the repulsive energy is inversely proportional to r^{12}. The total energy of van der Waals interactions of atoms, molecules, or groups may be represented by the following formula:

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \qquad (3.11)$$

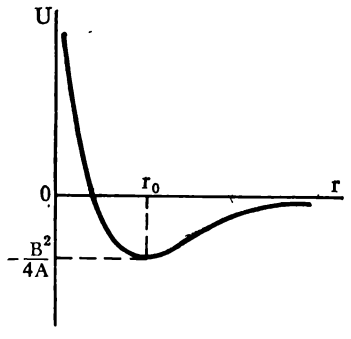


Fig. 3.3. The potential energy of van der Waals interaction versus the intermolecular distance.

This is the *Lennard-Jones (12-6)-potential*. Figure 3.3 shows the shape of the corresponding curve. The interaction energy is at a minimum and is equal to $-B^2/4A$ at a separation of $r_0 = (2A/B)^{1/6}$. The potential (3.11) is empirical. Other potentials have also been proposed: in particular, the repulsion may be represented by the exponential function $A \exp(-ar)$. This introduces another empirical parameter, namely a.

Table 3.2 presents the Lennard-Jones potential parameters for selected pairs of atoms.

The calculations of the energy of internal rotation are based on the assumption that intramolecular nonbonded interactions are of the same nature as intermolecular interactions. For example, Kitaigorodsky worked out, for these purposes, a *method of atom-atom potentials*, which had been developed by him earlier for the calculation of the energies of molecular crystals. The empirical parameters

Table 3.2. Lennard-Jones (12-6)-Potential Parameters

Atoms	$B\times10^6$, kJ/(mole·nm^6)	$A\times10^{12}$, kJ/(mole·nm^{12})	r_0, nm
HH	197.4	18.9	0.24
CC	1554.0	1201.2	0.34
NN	1524.6	676.2	0.31
OO	1541.4	609.0	0.30
CO	1541.4	861.0	0.32
CN	1537.2	907.2	0.32
CH	537.6	159.6	0.29
NO	1533.0	642.6	0.31
HO	520.8	105.0	0.27
NH	525.0	113.4	0.28

used are the equilibrium distances between nonbonded atoms, r_0. Besides, it is taken into account that in molecules the potentials have no spherical symmetry; for example, the equilibrium radius of a monovalent atom depends on the angle formed by this radius and

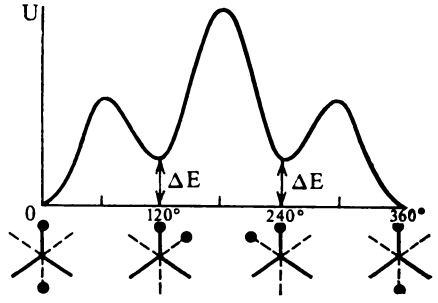
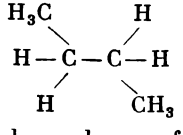


Fig. 3.4. The potential energy of internal rotation in *n*-butane as a function of the angle of rotation.

the valence bond. Thus, a further parameter is introduced. We shall return to the calculations of the conformations of molecules in the next chapter.

If a molecule, unlike ethane, has no axial symmetry, the $U(\varphi)$ curve can no longer be described by a formula of the type (3.9). For example, for the molecule of *n*-butane



the curve representing the dependence of the internal energy on the angle of rotation φ about the central C—C bond has the shape

shown in Fig. 3.4. There are three energy minima: the first is the lowest (at 0° and 360°) and corresponds to the *trans*-conformation (the staggered conformation) and the other two are of the same value and correspond to conformations produced by rotation from the *trans*-conformation of one C_2H_5 group relative to the other by 120° and −120°. These conformations are called *gauche* or *skew conformations.*

Evidently, molecules that are characterized by several nonequivalent energy minima, $U(\varphi)$, exist precisely in these states, changing from one conformation to another at a rate determined by the barrier height that separates these minima. The equilibrium numbers of *n*-butane molecules in the *trans* (N_t) and two gauche (rotated to the right and left by 120°) conformations (N_d and N_l) are readily calculated with the aid of the Boltzmann distribution law:

$$N_t = N\frac{1}{1+2\exp(-\Delta E/RT)}, \quad N_d = N_l = N\frac{\exp(-\Delta E/RT)}{1+2\exp(-\Delta E/RT)} \tag{3.12}$$

where ΔE is the difference between the energy minima (Fig. 3.4). Obviously, $N_t + N_d + N_l = N$ is the total number of molecules. The substance is a dynamic mixture of conformations, which are variously called *rotational isomers*, *rotamers*, or *conformers*. The composition of a thermodynamically equilibrium mixture is determined by the energy difference between the rotamers, ΔE, and by temperature. At $T \to \infty$ the number of molecules $N_t = N_d = N_l = N/3$; when T drops to the melting point the substance crystallizes in the form of the most stable rotamer and formulas (3.12) are no longer applicable.

At potential barrier heights of the order of 10 kJ/mole the time of rotational isomerization, i.e., the time required for one rotamer to be converted to another, is of the order of 10^{-10} s. This estimate is obtained through calculations based on the theory of absolute reaction rates (see Sec. 6.1). Hence, rotamers cannot be separated. Their presence and fraction are found by studying the physical and chemical properties of a mixture of rotamers. The spatial structure of rotamers is varied; accordingly, their vibrational spectra are also different. Hundreds and thousands of vibrations (with frequencies of the order of 10^{12} to 10^{13} s^{-1}) occur during the lifetime of the rotamer—the rotamer has enough time to display its spectrum. As a matter of fact, the existence of rotational isomerism was first established by Kohlrausch with the aid of Raman spectra. The ratio of the spectral line intensities, which correspond to different rotamers, depends on their content in the mixture in accordance with formulas (3.12). Hence, ΔE can be determined from the temperature course of the spectral line intensities. It has been found, for example, that for *n*-butane $\Delta E \approx 2.5$ kJ/mole.

The information on rotational isomerism can also be obtained with the aid of radiofrequency spectroscopy, NMR, electron diffraction studies, measurements of dipole moments, etc. A theoretical calculation of the values of ΔE is naturally carried out by the same methods that are employed for the calculation of potential energy barriers (page 101). If the energy of the system depends on several angles (in the case of *n*-butane, on the angles

$$\varphi(-H_2C\rightarrow CH_2-) \text{ and } \varphi_1=\varphi_2(H_3C\rightarrow CH_2-)],$$

the multidimensional energy surface $U(\varphi, \varphi_1, \varphi_2, \ldots)$ is calculated rather than a curve of the type shown in Fig. 3.4. This energy surface is a kind of a contour map. Such potential energy surfaces will be described in discussing the conformations of proteins.

It should be stressed that there is a fundamental difference between the *conformations* and *configurations* of molecules. The change from one conformation to another occurs as a result of internal rotations about single bonds. The configuration is not changed in this case. On the contrary, for the configuration to be changed, it is necessary that the covalent bonds be broken and new ones be formed. An example of different configurations is provided by the optical antipodes of chiral molecules (Fig. 2.10, page 46).

The study of rotational isomerism and isomerization, i.e., conformations and conformational changes, has become very important in organic and bioorganic chemistry. The main features of the physical behaviour of synthetic and biological macromolecules are associated with their conformational properties.

3.3. The Rotational-Isomeric Theory of Macromolecules

The polyethylene chain (page 63), which consists of single C—C bonds, undergoes internal rotations in each link of the chain. It is precisely the internal rotation, i.e., conformational changes, that determines the chain flexibility, which is responsible for the rubber elasticity of polymers.

Let us construct a maximally simplified model of the macromolecule in the form of freely-jointed rods of the same length b (Fig. 3.5). Each rod rotates freely relative to the adjacent ones. Thereby, a set of conformations that arise from rotations around a given atom of the freely-jointed chain is continuous in the range of angles from 0 to 4π and the energy is not changed on rotations. The chain can be characterized by the vector $\boldsymbol{h}$ drawn from the first atom in the chain to the last one (Fig. 3.5). It is evident that the value of the vector $\boldsymbol{h}$ averaged over all the conformations is equal to zero:

$$\bar{\boldsymbol{h}} = 0$$

since on thermal motion all its directions are equally probable. One must find out in this case the distribution of the probabilities of realization of the various numerical values of $h \equiv |\boldsymbol{h}|$ which can vary from zero to the maximum length of the extended chain, Zb (Z is the number of links and b is the length of the link). The solution of this problem is analogous to the solution of the problem in the theory of Brownian motion, namely the finding of the probability of a particle being displaced to a distance h as a result of Z

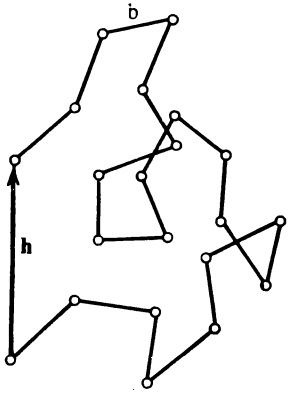


Fig. 3.5. The freely jointed chain.

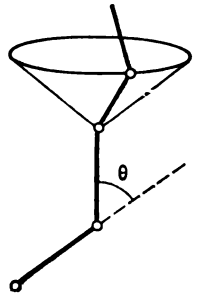


Fig. 3.6. The chain with fixed valence angles.

steps, each of which has a length b and an arbitrary direction. The distribution of the probabilities is found to be Gaussian—the probability that the distance between the chain ends lies in the interval between h and $h + dh$ is equal to

$$W(h)\,dh = \left(\frac{3}{2\pi Zb^2}\right)^{3/2} 4\pi h^2 \exp\left(-\frac{3h^2}{2Zb^2}\right) dh \qquad (3.13)$$

If we fix one end of the chain at point 0 and put the other end in a volume element at a distance h from 0, then the number of chain configurations will be proportional to the volume element and to the probability density $w(h)$ in this element. So we have

$$w(h) = \left(\frac{3}{2\pi Zb^2}\right)^{3/2} \exp\left(-\frac{3h^2}{2Zb^2}\right) \qquad (3.14)$$

The mean-square end-to-end distance of the chain is equal to

$$\overline{h}^2 \approx \int_0^\infty h^2 w(h)\,dh = Zb^2 \qquad (3.15)$$

that is, the chain is strongly coiled. Formula (3.15) is valid at $Z \gg 1$ when the upper limit of Zb can be replaced by infinity.

Thus, the thermal motion twists the macromolecule into a coil. The probability of such a state is much higher than that of the stretched state, since the extended chain conformation can be realized in a single way, while the coiled conformation can arise in a large variety of ways. The coil conformation of macromolecules in solution has been proved by direct experiments (Sec. 3.6).

The entropy nature of the high-elasticity of natural rubber follows directly from what has been said above. In the unextended state the chains are twisted into coils, to which there corresponds the maximum entropy. When the coil is extended, the entropy decreases. The entropy of the chain being stretched is given, according to Eq. (3.14) by the following formula (k is Boltzmann's constant):

$$S = \mathrm{k} \ln w(h) = C - \frac{3\mathrm{k}h^2}{2Zb^2} \tag{3.16}$$

and, in accordance with Eq. (3.8), the elastic force that arises on extension of the chain is equal to

$$f = -T\left(\frac{\partial S}{\partial h}\right)_T = \frac{3\mathrm{k}T}{Zb^2}\, h \tag{3.17}$$

Hooke's law is valid, the modulus of elasticity being proportional to absolute temperature. What we have described above is the kinetic theory of rubber elasticity proposed by Kuhn.

In real macromolecules, the valence angles between the bonds are fixed and rotations are not free. Let us specify the positions of two neighbouring links of the polyethylene chain (Fig. 3.6). The third link may occupy various positions on the surface of a cone with an aperture of 2θ ($\pi - \theta$ is the valence angle between the C—C bonds which is close to the tetrahedral angle, 109° 28′). To these different positions, which are characterized by the angle of rotation φ about the second C—C bond, there correspond different values of the potential energy $U(\varphi)$. The position of the fourth link relative to the first two is less definite, since it lies on the cone described around each of the positions of the third link, etc. A sufficiently remote link occupies a practically arbitrary position relative to the first link. Therefore, a long chain twists into a coil. The macromolecule can be split mentally into segments, whose positions are no longer correlated. Formulas (3.13) through (3.17) are still valid, but Z and b represent the number and length of freely-jointed segments rather than of real links.

It is possible, however, to construct a molecular theory of the chain, expressing its length through the number of real links, N, the length of the link l and the angle θ. The vector $\boldsymbol{h}$ is the sum of

the vectors of the links:

$$\boldsymbol{h} = \sum_{i=1}^{N} \boldsymbol{l}_i \tag{3.18}$$

Hence,

$$\overline{h^2} = \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{\boldsymbol{l}_i \boldsymbol{l}_j} = \sum_{i=1}^{N} l_i^2 + 2 \sum_{i=2}^{N} \sum_{j=1}^{i-1} \overline{\boldsymbol{l}_i \boldsymbol{l}_j} \tag{3.19}$$

In a real chain the mean values of the scalar products $\overline{(\boldsymbol{l}_i \boldsymbol{l}_j)}$ depend on the angles θ and φ. For a chain with symmetrical side groups, a chain of the polyethylene type, a strict calculation gives the *Oka formula* (at $N \gg 1$):

$$\overline{h^2} = Nl^2 \frac{1+\cos\theta}{1-\cos\theta} \frac{1+\eta}{1-\eta} \tag{3.20}$$

where η is the mean cosine of the angle of internal rotation φ and is defined by

$$\eta = \overline{\cos\varphi} = \frac{\int_0^{2\pi} \cos\varphi \exp(-U(\varphi)/kT)\, d\varphi}{\int_0^{2\pi} \exp(-U(\varphi)/kT)\, d\varphi} \tag{3.21}$$

The quantity $\overline{h^2}$ as calculated from Eq. (3.20) is greater than the value obtained for the freely-jointed chain (Nl^2), since the correlation determined by the absence of free rotation and the valence angles extends the chain. The quantity $\overline{h^2}$ is a measure of the *thermodynamic flexibility*—the lower the value of $\overline{h^2}$ at given N and l, the greater is the extent to which the chain is twisted and the greater is its flexibility. The calculation of $\overline{h^2}$ reduces to the calculation of η, which requires a knowledge of the potential energy $U(\varphi)$. This difficulty can, however, be avoided.

The theory of the sizes, shapes, and electrical and optical properties of macromolecules in solution must be based on the physical mechanism of their flexibility. This mechanism is disclosed with the aid of the *rotational-isomeric theory* (Volkenshtein, 1951-1958).

As we have seen earlier, small molecules devoid of axial symmetry form, as a result of rotations about single bonds, various rotational isomers (rotamers or conformers). Figure 3.7 shows the rotamers of *n*-butane (cf. Fig. 3.4); Fig. 3.8 presents the similar rotamers of polyethylene which are produced by rotation about any C—C bond in the chain. The only difference is that polyethylene contains CH_2 groups instead of CH_3 groups; the CH_2 groups are joined to the chain, which has a continuation (shown by a wavy line). From this it follows that the macromolecule may be regarded

as an equilibrium mixture of rotational isomers, and the internal rotation in chain as rotational isomerization, i.e., transitions from one conformation to another. This means that, instead of the continuous curve of $U(\varphi)$, it is sufficient to consider the states of the chain that correspond to rotamers with discrete, fixed values of φ. The

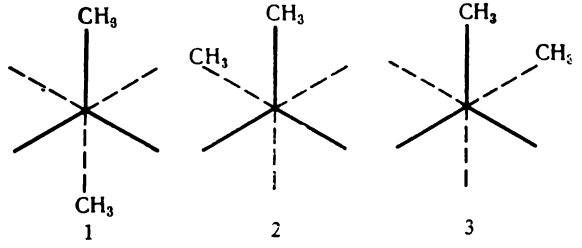


Fig. 3.7. The rotamers of *n*-butane.

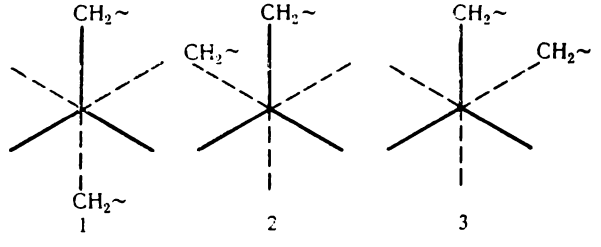


Fig. 3.8. The rotamers of polyethylene.

integration in Eq. (3.21) is replaced by summation over the rotamers. Thus, for polyethylene, which has three rotamers, *1*, *2*, and *3* (Fig. 3.8), we have

$$\eta = \frac{\cos\varphi_1 \cdot \exp(-E_1/kT) + \cos\varphi_2 \cdot \exp(-E_2/kT) + \cos\varphi_3 \cdot \exp(-E_3/kT)}{\exp(-E_1/kT) + \exp(-E_2/kT) + \exp(-E_3/kT)} \tag{3.22}$$

For polyethylene $E_2 - E_1 = E_3 - E_1 = \Delta E$, $\varphi_1 = 0°$, $\varphi_2 = 120°$, and $\varphi_3 = -120°$ (cf. Fig. 3.4):

$$\eta = \frac{1 - \exp(-\Delta E/kT)}{1 + 2\exp(-\Delta E/kT)} \tag{3.23}$$

At $\Delta E = 2500$ J/mole (page 70) and $T = 273$ K, the value of η is equal to 0.37. Using the Oka formula (3.20) with $\cos\theta =$

$= \cos(\pi - 109° 28') = -1/3$ we find that $\overline{h^2} = 1.1\, Nl^2$. The thermodynamic chain flexibility is the higher, the lower the value of η, i.e., the lower the difference ΔE.

Of course, in actual fact there occur fluctuations—torsional vibrations about the values of φ that correspond to rotational isomers. These may, however, be ignored, since these vibrations, being accidental, cancel each other on the average and do not affect the averaged properties of the chains.

The rotational-isomeric theory has been well substantiated for those cases where the minima of the energy $U(\varphi)$ are separated by barriers which by far exceed kT. If the barriers are low, the rotational-isomeric theory is still applicable as an approximate mathematical method, in which integration is replaced by summation.

The actual existence of rotamers in macromolecules has been established by a number of methods, primarily by the method of infrared spectroscopy in studying the thermomechanical properties and the extension of polymers. The rotational-isomeric theory forms the basis for the statistical physics of macromolecules. It makes possible the calculation not only of the sizes of macromolecular coils but also of their dipole moments and polarizabilities which are responsible for the electrical and optical properties.

The calculation of molecular flexibility is based on the chemical structure of macromolecules. So far we have spoken of polyethylene only. Many other macromolecules, however, contain in their links bulky side groups, say, polystyrene $(-CH_2-CHR-)_n$, where R is C_6H_5. In such cases the conformations are predominantly determined by the interactions between side groups. Information on chain conformations can be obtained by X-ray structural analysis provided that the polymer is capable of crystallization. Upon crystallization distinct rotamers become fixed for all chain links and a long-range order appears: knowing the positions of atoms in a given monomer unit, we can identify their positions in any other link, however remote they are, since the arrangement of atoms is strictly periodic. At the same time, there is also a short-range order in a crystal, of course; this is a certain definite arrangement of neighbouring links. The crystalline short-range order is retained on melting and dissolution of a polymer since the crystalline structure of the polymer corresponds to the minimum of the potential energy. It may be assumed that the one-dimensional short-range order in a free macromolecule, which forms a random coil, is analogous to a one-dimensional long-range order in a crystal. This idea has been verified in calculations of conformations and in the results of experimental investigations.

Rotations in neighbouring links are not independent of one another; there is a correlation which lengthens the chain. In polyethylene, the conformations of neighbouring links $(0°, 0°)$ are stable, i.e.,

trans-trans (*tt*); (0°, 120°) and (0°, −120°) (*ts*); (120°, 0°), (−120°, 0°) (*st*); (120°, 120°) and (−120°, −120°) (*ss*). The conformations (120°, −120°) and (−120°, 120°), i.e., (s^+s^-) and (s^-s^+), are energetically unfavourable. The mean-square length of the chain calculated with account taken of the correlation is found to be equal to $1.6Nl^2$ rather than to $1.1Nl^2$.

For an unsymmetric chain $(-CH_2-CHR-)_n$ the rotations in clockwise and counterclockwise directions are not equivalent and

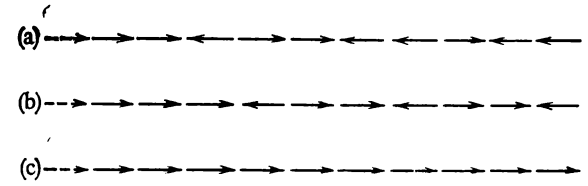


Fig. 3.9. A one-dimensional model illustrating the extension of the chain.

formula (3.20) does not hold. The expression for $\overline{h^2}$ contains $\overline{\sin \varphi}$ in addition to $\overline{\cos \varphi}$. In a symmetrical case $\overline{\sin \varphi} = 0$.

The rotational-isomeric theory offers a quantitative interpretation of the physical properties of macromolecules in solution which is in good agreement with experiment. It discloses the physical mechanism of the extension of natural rubber which consists in rotational isomerization, conformational transitions. This can be illustrated by a one-dimensional model of the macromolecule. Let us denote a link by an arrow of length l, which is directed to the right or left. One rotamer (designated as t) is represented by two adjacent arrows pointing in one direction, and the other (s) by two arrows pointing in opposite directions. The total length of the chain is expressed by the vector sum of all the lengths of the arrows. Figure 3.9*a* shows a chain consisting of 10 links, five of which are directed to the right and five to the left; hence the total chain length is $h = 0$. The chain contains five t-rotamers and five s-rotamers (so that the number of rotamers is equal to the number of links, i.e., 10; there is also added an additional, "zeroth", link directed to the right—it is marked by a dotted line). We stretch the chain by applying a force which turns the arrows to the right. Figure 3.9*b* shows one of the extended states of the chain having a length of $h = (7-3)\ l = 4l$. The number of rotamers remains the same, five t-rotamers and five s-rotamers. The chain is elongated up to $4l$ as a result of the redistribution of the rotamers, their relative content remaining unchanged. The transition *ttssts tsst* → *ttts tssts* has taken place, which is not accompanied by a change in energy. The chain cannot, however, be fully extended

by way of the redistribution of the rotamers without changing their number. The fully extended chain (Fig. 3.9*c*) consists of 10 *t*-rotamers. In this case, the conversion involves a change in both entropy and energy: $10E_t \neq 5E_t + 5E_s$. In fact, the mutual independence of the internal energy of rubber and its extension [formula (3.6)] is not fulfilled quite strictly. Along with the entropy contribution, there is observed a certain energy contribution to rubberlike elasticity, since the energy difference between the rotamers, ΔE, is nonzero. Therefore $(\partial E/\partial h)_T \neq 0$. This is confirmed by experiment. The investigations of the thermomechanical properties of rubbers have not only verified the presence of a slight energy elasticity but also made it possible to find experimental values of ΔE, which are in good accord with theoretical values (Flory). A direct study of the extension of polymers by means of infrared spectroscopy has shown that the relative content of rotamers is really changed on extension.

If the flexibility of the macromolecule is low (i.e., if the values of ΔE are high), there can be realized the mechanism of winding of the chain into a coil different from the rotational-isomeric scheme. Instead of the rotamerization, there occur torsional vibrations about a single value of φ. A very loose coil is produced. The double helices of DNA belong to the category of such rigid macromolecules. A stiff chain in which torsional vibrations take place may be regarded as an elastic rod (see Sec. 3.5).

So far we have discussed the thermodynamic equilibrium flexibility of macromolecules determined by the quantity ΔE. The *kinetic flexibility*, which characterizes the behaviour of the macromolecule in time, i.e., the rate of a conformational change, depends mainly not on ΔE but on the heights of the potential energy barriers that separate different conformations. Both the thermodynamic and kinetic flexibilities vary within wide limits. Macromolecules built up of conjugated π-bonds or of conjugated aromatic rings are rigid—they are devoid of conformational mobility. In distinction from macromolecules with σ-bonds (polyethylene, natural rubber), the conjugated chains are a kind of π-electronic semiconductors. Associated with this is the absorption of light in the long-wave region—such polymers are practically black in colour. The flexibility and semiconducting properties of macromolecules are incompatible (see page 153).

3.4. The Macromolecule as a Cooperative System

As has already been said and shown in Fig. 3.9, the state (conformation) of a given link of a macromolecule depends on the states (conformations) of the adjacent links. A system in which the states of the elements depend on one another is called cooperative. The concept of cooperativity was first introduced by Fowler in studying

phase transitions. These very important phenomena cannot be understood without reference to interactions between the elements of the system. For example, the gas-liquid transition is a cooperative transition. The equation that relates the gaseous and liquid states is the van der Waals equation of state:

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT \tag{3.24}$$

where a, b, and R are constants.

Let us compress a gas at a temperature below the critical temperature. A decrease in the volume V enhances the mutual attraction of molecules (a/V^2) which in turn causes a decrease in volume, etc. The cooperation of a large number of particles leads to a sharp transition, which proceeds according to the "all-or-none" principle.

Statistical mechanics has at its disposal methods for the calculation of the properties of cooperative systems: calculations of the partition function which is used to find all thermodynamic characteristics of the system. If the interrelation of conformations in a one-dimensional cooperative system, in a macromolecule, is taken into account, the expression for the potential energy will contain terms that are dependent on conformations, at least on the conformations of two neighbouring links:

$$U(\Omega_1, \Omega_2, \ldots, \Omega_N) = \sum_{i=1}^{N} U(\Omega_{i-1}, \Omega_i) \tag{3.25}$$

where Ω_i is the conformation of the ith link represented by the corresponding angles of rotation.

Below we give a calculation based on the *Ising one-dimensional model* and carried out with the aid of the matrix method proposed by Kramers and Vannier.

The conformational partition function for a macromolecule in the absence of external forces has the form

$$Z = \sum_{\Omega_1} \sum_{\Omega_2} \cdots \sum_{\Omega_N} \exp\left(-\frac{U(\Omega_1, \Omega_2, \ldots, \Omega_N)}{kT}\right) \tag{3.26}$$

The summation is carried out over r rotamers of each link. If expression (3.25) is valid, then

$$Z = \sum_{\Omega_1} \sum_{\Omega_2} \cdots \sum_{\Omega_N} \prod_{i=1}^{N} g(\Omega_{i-1}, \Omega_i) \tag{3.27}$$

where

$$g(\Omega_{i-1}, \Omega_i) = \exp\left(-\frac{U(\Omega_{i-1}, \Omega_i)}{kT}\right) \tag{3.28}$$

Considering that the chain consists of recurring identical units, we may regard values of $g(\Omega_{i-1}, \Omega_i)$ as the elements of a matrix:

$$\boldsymbol{G} = \begin{pmatrix} g(\Omega^{(1)}, \Omega^{(1)}) & \dots & g(\Omega^{(1)}, \Omega^{(r)}) \\ \vdots & & \vdots \\ g(\Omega^{(r)}, \Omega^{(1)}) & \dots & g(\Omega^{(r)}, \Omega^{(r)}) \end{pmatrix} \tag{3.29}$$

Let us introduce the notation

$$g(\Omega_{i-1}^{(m)}, \Omega_i^{(n)}) = G_{\Omega_{i-1}^{(m)}, \Omega_i^{(n)}} \tag{3.30}$$

The partition function (3.27) is rewritten in the following form:

$$Z = \sum_{\Omega_1} \sum_{\Omega_2} \dots \sum_{\Omega_N} G_{\Omega_0\Omega_1} G_{\Omega_1\Omega_2} \dots G_{\Omega_{N-1}\Omega_N} \tag{3.31}$$

and, according to the rule of matrix multiplication 'row by column', we find

$$Z = \sum_{\Omega_N} \boldsymbol{G}^N_{\Omega_0\Omega_N} \tag{3.32}$$

Introducing the cyclicity condition $\Omega_0 = \Omega_N$ (at $N \gg 1$ this condition plays no role), we obtain Z in the form of the trace Sp $\boldsymbol{G}^N$, i.e., the sum of diagonal elements of the matrix $\boldsymbol{G}$ raised to the power N:

$$Z = \mathrm{Sp}\, \boldsymbol{G}^N = \lambda_1^N + \lambda_2^N + \dots + \lambda_r^N \tag{3.33}$$

where λ_n are the eigenvalues of the matrix $\boldsymbol{G}$. All the elements of the matrix are positive. From this it follows that it has a maximum eigenvalue λ_1, which is real, positive and nondegenerate (the Frobenius theorem). At $N \gg 1$

$$Z \approx \lambda_1^N \tag{3.34}$$

Thus, the calculation of the rotameric partition function is reduced to finding the maximum root of the matrix λ_1. If the value of Z is known, we can calculate the equilibrium characteristics of macromolecules. In particular, this method is employed to find the mean-square length of the coil, $\overline{h^2}$ [see formula (3.20) and formulas that take account of the correlation of rotamers].

Let us consider, with the aid of this method, the extension of a one-dimensional cooperative model (Fig. 3.9) by applying a force f. Let the energy of the t-rotamer, i.e., the energy of two adjacent parallel links, be equal to $-E$, and the energy of the s-rotamer, i.e., that of two adjacent antiparallel links, be E. Hence, the energy difference between the rotamers will be

$$\Delta E = 2E \tag{3.35}$$

We write this condition in the following form:

$$E_{ij}=\begin{cases} -E\sigma_i\sigma_j & \text{if } i=j+1 \\ 0 & \text{if } i\neq j+1 \end{cases} \tag{3.36}$$

Here $\sigma_i = 1$ if the link (an arrow) is directed to the right and $\sigma_i = -1$ if it is directed to the left; i and j are the numbers of links. The energy $E_{ij} = 0$ if the links are not adjacent to each other—the correlation extends only to the neighbouring arrows. If an external force is applied, which orients the arrows in the right direction, then each of the links of length l acquires an extra energy:

$$E_f = -lf\cos\varphi = -lf\sigma_i \tag{3.37}$$

where φ is the angle between the direction of the link and that of the external force. Since each of the arrows may be only in two positions, then the matrix G has a second rank ($r = 2$). According to Eq. (3.28), we find

$$g(\sigma_j,\ \sigma_{j+1}) = \exp\left(\frac{E}{kT}\sigma_j, \sigma_{j+1} + \frac{lf}{kT}\sigma_j\right) \tag{3.38}$$

and the matrix G is obtained by substituting the values of σ_j and σ_{j+1} equal to $+1$ and -1:

$$G=\begin{pmatrix} g(1,\ 1) & g(1,\ -1) \\ g(-1,\ 1) & g(-1,\ -1) \end{pmatrix} \tag{3.39}$$

Introducing the notations $a = E/kT$ and $b = lf/kT$, we have

$$G=\begin{pmatrix} \exp(a+b) & \exp(-a+b) \\ \exp(-a-b) & \exp(a-b) \end{pmatrix} \tag{3.40}$$

The eigenvalues are found from the equation

$$\begin{vmatrix} \exp(a+b)-\lambda & \exp(-a+b) \\ \exp(-a-b) & \exp(a-b)-\lambda \end{vmatrix} = 0 \tag{3.41}$$

They are equal to

$$\lambda_{1,2} = e^a\cosh b \pm (e^{2a}\sinh^2 b + e^{-2a})^{1/2} \tag{3.42}$$

Since $\lambda_1 > \lambda_2$, then, according to Eq. (3.34),

$$Z \approx \lambda_1^N = [e^a\cosh b + (e^{2a}\sinh^2 b + e^{-2a})^{1/2}]^N \tag{3.43}$$

We have obtained a simple analytical expression for the partition function. Using this expression, we compute the average length of the chain:

$$\bar{h} = kT\frac{\partial \ln Z}{\partial f} = NkT\frac{\partial \ln \lambda_1}{\partial f} = Nl\frac{\sinh b}{(\sinh^2 b + e^{-4a})^{1/2}} \tag{3.44}$$

The measure of cooperativity is the quantity a, i.e., $\Delta E/\mathrm{k}T$. In the absence of cooperativity, i.e., at $a = 0$,

$$\bar{h} = Nl \tanh b = Nl \tanh \frac{fl}{\mathrm{k}T} \tag{3.45}$$

With the free orientation of the arrows the chain length is the larger, the greater is the force and the lower the temperature. At small forces, i.e., at $fl \ll \mathrm{k}T$,

$$\bar{h} \approx Nl^2 f/\mathrm{k}T \tag{3.46}$$

or

$$f \approx \frac{\mathrm{k}T}{Nl^2} \bar{h} \tag{3.47}$$

that is, the chain length is proportional to the tensile force (Hooke's law), the modulus of elasticity being proportional to absolute temperature. If, on the contrary, $a \gg 1$, i.e., $\Delta E \gg \mathrm{k}T$, then the quantity e^{-4a} in expression (3.44) may be neglected and $\bar{h} = Nl$, irrespective of the force. The chain is rigid, and therefore it is extended. The same result obtains with large forces, i.e., at $fl \gg \mathrm{k}T$. Here too $\sinh^2 b \gg e^{-4a}$—a large force fully extends the chain.

The root-mean-square length of the rotameric one-dimensional chain is calculated, in the absence of an external force, directly from formula (3.19):

$$\bar{h}^2 = Nl^2 + 2 \sum_{i=2}^{N} \sum_{j=1}^{i-1} (\overline{\boldsymbol{l}_i \boldsymbol{l}_j}) = Nl^2 \frac{1+\eta}{1-\eta} + 2l^2 \frac{\eta(\eta^N - 1)}{(1-\eta)^2} \tag{3.48}$$

where η is the mean cosine of the angle between two neighbouring arrows. In the case under consideration,

$$\eta = \frac{\cos 0^\circ \exp(E/\mathrm{k}T) + \cos 180^\circ \exp(-E/\mathrm{k}T)}{\exp(E/\mathrm{k}T) + \exp(-E/\mathrm{k}T)} = \tanh \frac{E}{\mathrm{k}T} \tag{3.49}$$

and at $N \gg 1$

$$\bar{h}^2 = Nl^2 \exp(2E/\mathrm{k}T)$$

The correlation determined by the advantage of the parallel arrangement of the arrows extends the chain.

The conformational transitions of macromolecules both on melting of a crystalline polymer and on extension of the chain are cooperative. This is important for the understanding of the properties of biopolymers.

3.5. The Coil and the Globule

The primary or chemical structure of the macromolecular chain is fixed—the chemical bonds in it do not exchange places. This means that the equilibrium state of the chain is, in fact, the state of

the partial equilibrium with fixed linear memory (Lifshitz). The presence of linear memory is the factor responsible for certain features in the properties and structure of macromolecules.

We have already considered the conformational changes of macromolecules, i.e., the rotations about single bonds. As a result of such rotations, a coil is formed, which is the looser, i.e., its size is the larger, the less flexible is the chain. For sufficiently rigid chains, apart from the model considered above, use may be made of the elastic rod model. Suppose that the chain consists of $N + 1$ links of length l, which form successive small angles $\varphi_1, \varphi_2, \ldots$ $\ldots, \varphi_N$ with one another. The cosine of the angle between the ith and $(i + j)$th links is given by

$$\cos \varphi_{i,\, i+j} = \frac{(l_i,\, l_{i+j})}{l^2} = \overline{(\cos \varphi)^j} \tag{3.50}$$

where $\overline{\cos \varphi}$ is the mean value of $\cos \varphi$. At small φ

$$\cos \varphi_{i,\, i+j} \approx (1 - \overline{\varphi^2}/2)^j \approx \exp\left(-j\overline{\varphi^2}/2\right) \tag{3.51}$$

and

$$\cos \varphi_{1,\, N+1} \approx \exp\left(-N\overline{\varphi^2}/2\right) = \exp\left(-L/d\right) \tag{3.52}$$

where $L = Nl$ is the contour chain length, and the quantity

$$d = \frac{2l}{\overline{\varphi^2}} \tag{3.53}$$

is the so-called *persistent chain length*, a parameter which characterizes the loss of correlation between the links during the motion along the chain. The meaning of the persistent chain length is as follows: a chain segment, shorter than d, may be regarded as a rigid rod, the segments of length d being capable of rotation independently of one another—they are freely-jointed. Theory gives the following value of the mean-square chain length:

$$\overline{h^2} = 2d^2\left(\frac{L}{d} - 1 + e^{-L/d}\right) \tag{3.54}$$

At $L \ll d$ the square end-to-end chain distance, $\overline{h^2}$, is approximately equal to L^2 and the chain is extended; at $L \gg d$, the value of $\overline{h^2}$ is equal to $2dL$ and the chain is coiled up with a statistical chain segment of length $2d$. Even a very rigid chain is coiled up if it is long enough.

Obviously, the links in the coil which are far away from one another along the chain can come closer together. So-called *volume effects* thus arise (Flory). The interactions of monomeric units in the coil are similar to the interactions of molecules in a real gas: a strong repulsion at small distances, as a result of which two monomers cannot be in the same place, and a weak attraction at large distances·

Simple calculations of the coil size do not take into account these effects. Nevertheless, such calculations may be compared with experiment at the so-called Flory Θ-point, which is defined as the temperature at which the attractive forces in a given solvent are compensated for by the repulsive forces and no volume effects are present. The theta temperature is similar to the Boyle point of a real gas. At $T > \Theta$ the repulsion predominates (a good solvent), and at $T < \Theta$ there predominates attraction (a poor solvent). The coil is dilated or contracted.

A macromolecular coil as a fluctuating system is characterized by a density correlation, i.e., a quantity which relates a change in density at one point in space occupied by the coil to a change in density at another point. It turns out that the correlation radius is of the same order as the size of the coil. The presence of a large correlation radius is due exactly to the linear memory in the chain. But if this is so, then the density of the coil is not its thermodynamic characteristic; it has no reliable constant value. In other words, the macromolecular coil is continuously fluctuating and its fluctuations are macroscopic. Only the time-averaged density of the coil has meaning.

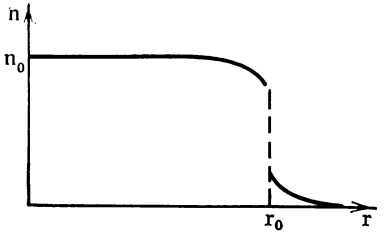


Fig. 3.10. The density distribution in the globule.

The Flory theory regards a polymer coil as a cloud of monomers, neglecting their union into a single chain. It is, however, necessary to deal not with monomer-monomer interactions but with interactions between chain portions. In other words, a polymeric coil may be depicted as a cloud of quasiparticles—quasi-monomers which represent the collective properties of all the monomers in the chain. The theory proposed by Khokhlov is rigorous and provides a good agreement with experiment.

If a strong attraction occurs between the chain links at $T < \Theta$ or the coil is acted on by an external compressing field, it is compacted into a tight coil structure, a *globule*. The process that takes place is similar to the condensation of a gas to a liquid.

The theory of the coil-globule transition has been worked out by Lifshitz, Grossberg, and Khokhlov. The phase transition coil → → globule and the state of the globule depend on the properties of the polymeric chain. If the chain is long and its flexibility low, i.e., if $\rho^3 \gg r^3$, where ρ^2 is the mean-square distance between neighbouring monomers in the coil, and r is a quantity which is of the

order of the radius of the interaction between chain links, then at $T < \Theta$ the phase transition is a transition of the first kind (first-order transition) with a jump in density. If the chain is flexible, i.e., $\rho^3 \approx r^3$, there occurs a smooth phase transition of the second kind (second-order phase transition) accompanied by a gradual swelling of the globule up to the size of the coil. At $\rho^3 \gg r^3$ the globule itself is a two-phase system, consisting of a compact core and a fluctuating "fringe", whose density falls off gradually to zero.

Figure 3.10 shows the density distribution in such a two-phase globule. As the temperature falls the density in the fringe declines just as the density of a vapour over a liquid drop. Because of the interconnection of the links in the chain at a sufficiently low temperature a short chain forms a fringeless globule. The core of the globule is similar to a crystal rather than to a liquid; it is characterized by zero configurational entropy. The formation of a globule means that it is the most preferred configuration of the links.

It should be stressed that this theory is at present restricted to the consideration of a homopolymer chain built up of identical monomeric units.

3.6. Methods of Investigation of Macromolecules

Macromolecules of any chemical structure are characterized by a definite molecular weight or by a molecular-weight distribution and, when being in the form of a coil or globule, also by their size and shape. Below we shall consider methods of experimental determination of these characteristics.

The most important molecular-weight determination methods employed in the physics of polymers are *osmometry*, *viscosimetry*, and *sedimentation in the ultracentrifuge*. The last method is especially widely used for the study of biopolymers.

Osmometry is based on the measurement of the osmotic pressure p_{osm} of a polymer solution. According to the van't Hoff law,

$$p_{osm}/c = RT/M \tag{3.55}$$

where c is the solution concentration in g/cm^3; M is the molecular weight (for polydisperse polymers M is the number-average molecular weight). Equation (3.55) holds for ideal solutions. For nonideal solutions (polymer solutions are invariably nonideal)

$$p_{osm} = \frac{RT}{M} c + Bc^2 + Cc^3 + \ldots \tag{3.56}$$

The coefficients B, C, etc. are the second, third, etc. *virial coefficients*. For dilute solutions the third and succeeding terms may be neglected. A plot of p_{osm} against c is linear and allows us to find M and B.

The *intrinsic viscosity* of a polymer solution is defined as

$$[\eta] = \lim_{c \to 0} \frac{\eta - \eta_0}{\eta_0 c} \tag{3.57}$$

where η is the viscosity of the solution; η_0 is the viscosity of the pure solvent. The quantity $[\eta]$ depends on M. It has been found that for a solution of polymer coils

$$[\eta] = AM^a \tag{3.58}$$

It should be noted that $0.5 \leqslant a \leqslant 1.0$, depending on the permeability of the coil to the solvent. For a fully permeable coil $a = 1.0$, and for a nonpermeable coil $a = 0.5$. For rigid spheres $a = 0$, and for long rigid rods $a = 1.7$. Formula (3.58) can be used for the determination of molecular weights if A and a are known from independent data.

The sedimentation method consists of the sedimentation of macromolecules under the action of a centrifugal force in the ultracentrifuge, whose rotor revolves at a speed of up to 10^4-10^5 rpm. The centrifugal acceleration is much greater than the force of gravity g—in modern ultracentrifuges it reaches 350,000 g (70,000 revolutions per minute). The cell containing a polymer solution is a cylinder with transparent windows made of crystalline quartz; it is placed in the rotor. The observation is carried out by optical methods.

If the rate of sedimentation considerably exceeds the rate of diffusion of macromolecules, sedimentation occurs. Two regions are formed in the cell—the pure solvent and the polymer solution. There is also a transient zone, where the solute concentration c varies from zero to a certain maximum values. As the sedimentation progresses this zone and its boundary are displaced. The velocity of travel of the boundary towards the base of the cell, i.e., from the axis of rotation of the rotor, is measured.

The macromolecule is acted on by a centrifugal force $V_m \rho_m \omega^2 x$, where V_m is the volume of the macromolecule; ρ_m is its density; ω is the angular velocity of the rotor; x is the distance from the axis of rotation. But in solution the macromolecule is acted on by the buoyancy, and the effective centrifugal force is equal to $V_m (\rho_m - \rho_0) \omega^2 x$, where ρ_0 is the density of the solvent. This force is balanced against the frictional force during translational motion, $\varkappa dx/dt$ (where $\varkappa$ is the frictional coefficient). When calculated per 1 mole we obtain:

$$N_A V_m \rho_m (1 - \overline{V}_m \rho_0) \omega^2 x = \varkappa' dx/dt \tag{3.59}$$

where N_A is Avogadro's number and $\overline{V}_m = \rho_m^{-1}$ is the specific partial volume of the macromolecule. But, on the other hand,

$$N_A V_m \rho_m = M \tag{3.60}$$

For dilute solutions $\varkappa$ is expressed in terms of the diffusion coefficient D according to the Einstein formula:

$$\varkappa = \frac{RT}{D} \tag{3.61}$$

Equations (3.59) through (3.62) lead to the Svedberg formula:

$$M = \frac{RTs}{(1-\bar{V}_m\rho_0)\,D} \tag{3.62}$$

where s is the *sedimentation coefficient* defined by

$$s = \frac{1}{\omega^2 x}\frac{dx}{dt} \equiv \frac{1}{\omega^2}\frac{d\ln x}{dt} \tag{3.63}$$

A sedimentation coefficient of the value of 1×10^{-13} s is called a Svedberg unit (1 S).

The positions at successive times of the sedimenting boundary layer are photographed in ultraviolet light. By measuring s, D, ρ_m, and ρ_0 it is possible to determine M. The sedimentation coefficient s depends on the solution concentration due to the hydrodynamic interaction of the molecules. The value of s extrapolated to zero concentration

$$s_0 = \lim_{c\to 0} s \tag{3.64}$$

is called the *sedimentation constant*. It increases with increasing M. For example, the values of s_0 for myoglobin, hemoglobin, and phycoerythrin are, respectively, 2.0, 4.5, and 12.0 S. The corresponding values of molecular weights are 17,600, 68,000, and 290,000.

With low rotational velocities of the ultracentrifuge a stage is reached where the rate of sedimentation equals the rate of diffusion and a state of equilibrium sedimentation occurs. At sufficiently large D the equilibrium sets in rapidly and there is observed an equilibrium distribution of the material.

If c is the concentration of the dissolved polymer at a distance x from the axis of rotation, then the number of molecules undergoing sedimentation during time dt through unit surface perpendicular to x is equal to

$$c\frac{dx}{dt}dt$$

where dx/dt is the rate of sedimentation at the distance x. During the same time dt the amount of the solute that diffuses through the same cross-sectional area is equal to

$$D\frac{dc}{dx}dt$$

At equilibrium

$$c\frac{dx}{dt} = D\frac{dc}{dx} \tag{3.65}$$

But, according to Eqs. (3.62) and (3.63),

$$\frac{dx}{dt} = s\omega^2 x = \frac{M(1-\bar{V}_m\rho_0)}{RT} D\omega^2 x \tag{3.66}$$

Substituting Eq. (3.66) into Eq. (3.65), we obtain the second Svedberg equation, which no longer contains the diffusion coefficient D:

$$M = \frac{RT}{(1-\bar{V}_m\rho_0)\,\omega^2 x} \frac{d\ln c}{dx} \tag{3.67}$$

The integration of this equation over the interval between x_1 and x_2 gives:

$$M = \frac{2RT}{(1-\bar{V}_m\rho_0)\,\omega^2} \frac{\ln(c_2/c_1)}{x_2^2 - x_1^2} \tag{3.68}$$

In the Archibald method, which requires a lower consumption of material and time, not the equilibrium itself but the approach to equilibrium is investigated. The flows of material through the meniscus and the base of the cell have a zero value, independently of the attainment of equilibrium. From Eqs. (3.65) and (3.66) it follows that

$$\frac{1}{x_m}\frac{1}{c_m}\left(\frac{dc}{dx}\right)_m = \frac{1}{x_0}\frac{1}{c_0}\left(\frac{dc}{dx}\right)_0 = \frac{s\omega^2}{D} \tag{3.69}$$

The subscript m refers to the meniscus and 0 to the bottom of the cell. Combining Eqs. (3.69) and (3.66) leads to the following expressions for molecular weight:

$$\begin{aligned} M &= \frac{RT}{(1-\bar{V}_m\rho_0)\,\omega^2} \frac{(dc/dx)_m}{x_m c_m} \\ M &= \frac{RT}{(1-\bar{V}_m\rho_0)\,\omega^2} \frac{(dc/dx)_0}{x_0 c_0} \end{aligned} \tag{3.70}$$

For a homogeneous polymer these two values of M coincide.

Of importance in biophysics is the *sedimentation in a density gradient*. In a concentrated solution of a low-molecular-weight compound (in CsCl, sucrose, etc.) on ultracentrifugation there is set up a concentration gradient, i.e., a density gradient of the solvent for macromolecules, $d\rho_0/dx$. In such a solution the macromolecules will take up a position in that part of the cell where $s = 0$, i.e., according to Eq. (3.66), $\bar{V}_m\rho_0 = 1$ or $\rho_0 = \rho_m$. In other words, the macromolecules will be accumulated in the part of the cell where the density of the concentrated solution coincides with the density of the macromolecules (the value of ρ_m is measured directly). A heterogeneous mixture of macromolecules is separated and there is observed a density spectrum. This method is efficiently employed in studying nucleic acids.

The diffusion is described by Fick's equation:

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \cdot \tag{3.71}$$

The diffusion coefficient D is found with the aid of formula (3.71). The concentration gradient and the rate of its change are determined by optical methods. In a special cell the pure solvent is placed over the solution and the light refraction or interference is studied.

For real solutions with a sufficiently high concentration, in place of the Einstein formula (3.61), the following expression is valid [cf. (3.56)]:

$$D = \frac{1}{\varkappa} (RT + 2BMc + 3CMc^2 + \ldots) \tag{3.72}$$

that is, D depends on the concentration. The frictional coefficient $\varkappa$ depends on the shape of the macromolecule and its permeability to the solvent. For rigid spheres of radius r Stokes' law holds:

$$\varkappa_{\text{sph}} = 6\pi\eta_0 r \tag{3.73}$$

For nonspherical particles the frictional coefficients are different for the directions of motion that coincide with the longitudinal and lateral axes of the particle. The mean frictional coefficient for an ellipsoid particle is higher than for a spherical particle of the same volume. For long rod-like molecules the frictional coefficient and viscosity are especially high.

Information on the molecular weight and shape of macromolecules is provided by optical methods of investigation. The intensity of the light scattered by a polymer solution depends on M. As a result of scattering, the intensity of the light passing through the solution is attenuated according to the law

$$I = I_0 \exp(-\chi l) \tag{3.74}$$

where I_0 is the incident light intensity; l is the path length travelled by the light ray through the solution; χ is the turbidity coefficient. The theory of light scattering gives

$$\chi = \frac{8\pi}{3} HcM \tag{3.75}$$

where

$$H = \frac{4\pi^2 n_0^2}{N_A \lambda^4} \left(\frac{dn}{dc}\right)^2 \tag{3.76}$$

Here n_0 is the refractive index of the solvent and n is the refractive index of the solution; λ is the wavelength of the light; N_A is Avo-

gadro's number. The value of M is found by measuring the turbidity coefficient.

When light is scattered by macromolecules, whose size is commensurate with the light wavelength, the *Mie effect* is observed, which is a specific asymmetric scattering indicatrix (the angular dependence of the scattered light intensity): the intensity of the light scattered forward is higher than that of the light scattered backward. The Mie effect is determined by the interference of light waves with different phases scattered by various parts of the molecules.

Associated with the Mie effect is a curious episode in the history of biophysics. Certain absorption bands characteristic of hemoglobin in solution could not be observed in the blood spectrum for a long time. On this basis it was stated that hemoglobin in erythrocytes is in an altered state. Later it was found that the absorption of light by hemoglobin was masked by the Mie effect—the enhanced scattering of light forward by erythrocytes.

Debye calculated the scattering function $P(\vartheta)$, i.e., the ratio of the light intensity scattered at an angle ϑ to the intensity of the incident light:

$$P(\vartheta) = I_\vartheta / I_0 \tag{3.77}$$

for macromolecules of various shapes—for spheres, ellipsoids, and rods. The measurement of the light scattering function makes it possible to determine the shape of the macromolecule. The Zimm method allows one to measure the molecular weight and size of macromolecules simultaneously.

Highly valuable information of biopolymers is supplied by the study of the scattering of X-rays by their solutions (see Chapter 5).

The shape of macromolecules (and also their anisotropic polarizability) can also be determined by means of flow birefringence (*the Maxwell dynamooptical effect*). In flow birefringence two coaxial cylinders are employed, one of which is stationary and the other rotated; the polymer solution occupies the space between the cylinders. The inner cylinder, called the rotor, revolves about the common axis, carrying off the liquid. A velocity gradient is set up in it—the layer adjacent to the wall of the rotor moves most rapidly; the layer adjoining the wall of the stationary cylinder is immobile. As a result, the macromolecules become oriented in solution and are acted on by a tensile force. The liquid becomes anisotropic, like a diaxial crystal. Birefringence is observed in a direction parallel to the axis of the dynamooptimeter. Its measurement provides the information indicated above.

On the contrary, birefringence in an electric field (*the Kerr effect*) is of much less importance in the case of biopolymers. Biopolymers are polyelectrolytes: the macromolecules carry charges and therefore not only become oriented but also move in an electric field.

3.7. Polyelectrolytes

Many macromolecules, including proteins and nucleic acids, are macroions in aqueous solution, carrying a multitude of charged groups. An example is polymethacrylic acid

$$\left[\begin{array}{l} -C(CH_3)-CH_2- \\ \quad | \\ \quad COO^- \end{array}\right]_n$$

which is a polyanion. Biopolymers are polyampholytes, i.e., they contain both anionic and cationic groups. Proteins contain acidic (Glu, Asp, Tyr) and basic (Arg, His, Lys) residues (see page 28); nucleic acids contain acidic phosphate groups and the basic groups $N\equiv$, $HN=$, and H_2N- of nitrogenous bases.

There exist solid macroions (globular proteins, etc.) and flexible polyelectrolyte chains. Evidently, the properties of such macromolecules depend substantially on the interactions of the charges present in them. The magnitudes of these charges depend on the degree of dissociation of ionogenic groups and the ion atmosphere that surrounds them.

The conformation of a flexible polyelectrolyte chain is determined by the minimum condition for the sum of conformational and electrical free energies. Naturally, the presence of like charges in the chain means their mutual repulsion, which leads to the uncoiling of the macromolecule, whose size is thus increased. The electrostatic free energy of the coil is calculated with account taken of the ion atmosphere. Flory has worked out a theory of the sizes of polyelectrolyte chains which is similar to the theory of volume effects proposed also by him (page 83). It is assumed that the coil and the solvent immobilized by it are electrically neutral as a whole. Calculations show that electrostatic interactions cannot convert the coil into an extended chain—only the expansion of the coil occurs. This is in agreement with experimental data, with the dependence of intrinsic viscosity $[\eta]$ on molecular weight. In a more rigorous statistical theory of charged macromolecules it is taken into account that, because of the shielding by counterions, the charged groups of the macromolecule situated far away from one another along the chain interact only during their accidental approach due to fluctuations. From this theory it follows that the conformational properties of charged macromolecules are intermediate between the properties of unswollen random coils and rigid rods. For polyelectrolytes $\overline{h}^2 \sim M^{4/3}$ and $[\eta] \sim M$ (cf. page 86). These results agree with experiment (in particular, for denatured DNA). The square of the linear swelling coefficient is proportional to $I^{-2/3}$ (I is the ionic strength, page 31).

The greatest extent of uncoiling of the macromolecule of a polyelectrolyte is attained at a low ionic strength. Figure 3.11 shows the

effect of the charge on the radius of gyration of the macromolecule of polymethacrylic acid. These data have been obtained by the light scattering method for solutions containing no salts. The degree of dissociation of the polyacid is controlled by the addition of a base. In such a case, of course, the ionic strength does not remain constant.

The expansion of the coil increases with dilution of the solution, since a decrease in concentration of the polyelectrolyte at constant degree of dissociation means a decrease in ionic strength. Information on the structure and properties of polyelectrolyte macromolecules

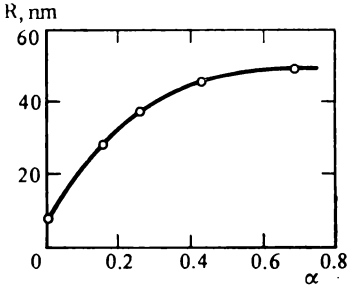


Fig. 3.11. The dependence of the radius of gyration of macromolecules of polymethacrylic acid on the degree of dissociation of the carboxyl groups.

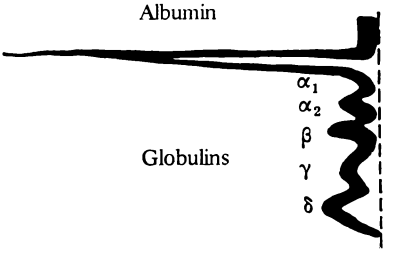


Fig. 3.12. An electrophoretic pattern.

in solution can be obtained through the use of isoionic dilution: the polyelectrolyte solution is diluted by a salt solution with the ionic strength equal to the ionic strength of the most concentrated polyelectrolyte solution in order to maintain the total concentration of counterions constant. Under these conditions $(\eta - \eta_0)/\eta_0 c$ is linearly dependent on c.

Counterions can be bound by the charged groups of a polyelectrolyte in a specific way—such binding depends on the chemical nature of the macroion and the small ion. This specific binding, which boils down to the formation of salt linkages at fixed points of the macromolecule, should be differentiated from nonspecific binding, which consists of the formation of an ionic atmosphere. The counterion in a salt linkage is at a considerably shorter distance from the charged group of the polyion than the distance between this group and mobile counterions. The specific binding of counterions determines the ion-exchange properties of polyelectrolytes which find impor-

tant practical applications. Cross-linked insoluble polyelectrolytes, which undergo swelling in water or in other liquid media, are used as ion-exchange resins or *ionites*. Ion-exchange resins are capable of sorbing certain ions from solutions; this ability of ion-exchange resins has found application for the purification and separation of various electrolytes and for the purification of nonelectrolytes from ion impurities.

In polyampholytes and, hence, in biopolymers, salt bridges (ionic bonds) can be formed between the cationic and anionic groups in a single chain or in several chains. Investigations of the structure and properties of biopolymers must, of necessity, take into account their polyampholyte nature and, hence, the pH and ionic strength of the medium. The structure of native (i.e., biologically functional) molecules of proteins and nucleic acids depends to a considerable extent on the electrostatic, ionic interactions. Of no less importance are interactions with small ions of the medium. The interaction between proteins and K^+, Na^+, Ca^{2+}, and Mg^{2+} ions determines the most significant biological phenomena, in particular the generation and transmission of the nervous impulse and muscle contraction. The functional structure of nucleic acids and their participation in protein biosynthesis are also associated with the cations of alkali and alkaline-earth metals.

The properties of polyelectrolytes allow us to simulate mechanochemical processes, particularly muscle contraction (see Chapter 12).

The polyelectrolyte properties of proteins make possible their separation and investigation by *electrophoresis*. If a solution containing charged particles is in an electric field, its particles move towards the respective electrodes. The electrophoretic mobility of a protein depends on the net charge of its molecules at given values of pH and ionic strength. Special electrophoretic methods have been devised for the analysis of proteins and their mixtures. The moving boundary method or frontal electrophoresis are concerned with the determination of the distance travelled in an electric field by the sharp boundary between the solution of the substance under study and the solvent. For preparative separation of protein mixtures, use is made of the method of zone electrophoresis. The solution of the compound to be studied is first placed as a narrow layer between two layers of the solvent. For a sufficiently long period of time compounds having different mobilities move apart to a considerable distance, their zones no longer overlap and the separated products can be extracted from the solution. The separation is hampered by the convection transport of the zones. Convection is reduced to a minimum, using a supporting porous medium—gels, paper filter, etc. To stabilize the system, use is successfully made of a density gradient (cf. page 88). Electrophoresis in this case is conducted in a vertical

tube containing a buffer solution whose density increases with the density gradient.

Just as in the case of sedimentation and diffusion, the recording of the electrophoretic pattern is made by means of optical methods. The resulting electrophoretic pattern (Fig. 3.12) consists of a series of peaks corresponding to the various components of the mixture. The peak area characterizes the amount of the corresponding component.

Chapter 4 Physics of Proteins

4.1. The Goals of Protein Physics

Proteins are the invariable participants in all life processes. Protein-enzymes catalyze all chemical, electrochemical, and mechanochemical processes taking place in cells and in organisms. The most important function of proteins is that of enzymes. Specialized enzymes serve as catalysts for all metabolic reactions, DNA replication, the transcription of the genetic message from a DNA strand to mRNA, and the transfer of the information encoded in the mRNA to the ultimate structure of the protein molecule (the process referred to as translation). Proteins also function as regulators of the genetic functions of nucleic acids. Regulatory enzymes, called *allosteric enzymes* (Chapter 6), provide feedback in metabolic chains.

The movement of cells and organisms, the performance of mechanical work by them (say, muscular work), is realized by special contractile proteins, which serve as working substances for these processes. Contractile proteins perform the enzymic, ATPase function, accomplish the conversion of the chemical energy (stored in ATP, page 43) to mechanical work. The "charging-up of the storage battery", i.e., oxidative phosphorylation, occurs in mitochondrial membranes with the requisite invariable participation of the respiratory chain enzymes. Oxidation-reduction enzymic processes also occur in photosynthesis. Other membrane proteins are responsible for the active transport of molecules and ions through the membranes and, consequently, for the generation and transmission of the nervous impulse. Proteins control all metabolic and bioenergetic processes.

In animal organisms there are special proteins which perform specific functions. The proteins serve to store oxygen (myoglobin) and transport it (hemoglobin, hemocyanin). Low-molecular-weight proteins or, more exactly, polypeptides, are hormones (page 54). The gamma-globulins of higher organisms protect them against foreign biopolymers by functioning as antibodies in immune processes. Finally, proteins that enter into the composition of connective tissues, cartilages and tendons and also the proteins of skin, hair,

and feathers are structural proteins which provide a reliable and mobile active intercommunication between the organs, the unity of the organism and its protection against external influences.

One can say that proteins are invariably involved in the storage, transfer, transformation, recoding, and reception of chemical signals—macromolecules, molecules, and ions—in living systems. In many cases, the proteins themselves serve as signals, their receptors, and transformers. Proteins that enter into the composition of the receptor systems of the organism recodes external signals into the chemical and electrochemical language.

Protein molecules are the most complex of the molecules known to science. Their biologically functional spatial structure and also the structure of supermolecular systems containing proteins (membranes, etc.) are determined both by chemical bonds in protein chains and by the entire set of weak interactions. Native proteins are never random coils. Globular proteins are aperiodic crystals of complex structure. They are dynamic rather than statistic systems, a sort of a machine, whose behaviour depends on the position and properties of all its component parts. Apart from globular proteins, there also exist fibrous proteins—contractile and structural proteins.

Protein denaturation, i.e., the loss of biological functionality by a protein when it is heated, acted on by acids and bases, etc., consists in the destruction of weak interactions and, as a result, in the final transition of the aperiodic crystal (globular or fibrous) into a random coil.

Let us formulate the principal goals of the physics of the protein.

1. The theoretical and experimental investigation of the structure of protein molecules and supermolecular systems that contain such molecules.

2. Directly related to these investigations is the problem of establishing the relationship between the primary structure of the protein chain and the three-dimensional structure of the native macromolecule.

3. The study of the physical mechanisms of formation of the primary structure of a protein, i.e., the mechanisms of genetic coding, translation of the genetic information, and of the regulation of the action of genes by proteins.

4. The study of physical mechanisms that control the various biological functions of proteins, primarily the enzyme activity.

The solution of these problems will lead to the understanding of the essence of biological phenomena in cells and organisms on the molecular, physical (and chemical) basis.

4.2. Conformations of the Polypeptide Chain

To understand the protein structure, it is necessary to consider the possible conformations of the polypeptide chain. These conformations are primarily determined by the structure of the peptide bond —CO—NH— (page 35). The structural parameters of peptide units established by means of the X-ray analysis of peptides and related compounds are given in Table 4.1. The meaning of the symbols C^{α} and C is clear from Fig. 4.1. The letters X and Y stand for atoms linked to C^{α} both in the backbone chain and in the side group R. The fully extended chain (without deformation of the valence angles and changes of the bond lengths) has a *trans*-conformation with zero values of the rotational angles φ, ψ, and ω (Fig. 4.1).

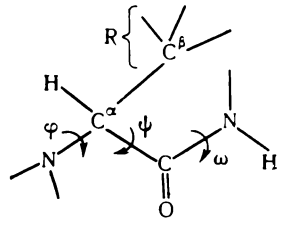


Fig. 4.1. A fragment of the protein chain.

Such a conformation is not, however, the most stable. The hydrogen atoms of the imine groups N—H form *hydrogen bonds* with the oxygen atoms of the carbonyl groups:

$$\rangle C{=}O \cdots H{-}N\langle$$

(see Sec. 4.3)

In order to find the most stable conformation of the polypeptide chain, it is necessary to minimize its total energy, including the energy of the intramolecular hydrogen bonds.

Pauling and Corey have found the most stable conformations of the polypeptide chain on the basis of the data obtained by X-ray

Table 4.1. The Structural Parameters of Peptide Units: Bond Lengths and Valence Angles

Bond	Bond length, nm	Valence angle, degrees	
$C^{\alpha}-C$	0.153	$C^{\alpha}CN$	114°
C—N	0.132	OCN	125°
$N-C^{\alpha}$	0.147	$C^{\alpha}CO$	121°
C=O	0.124	CNC^{α}	123°
N—H	0.100	CNH	123°
$C^{\alpha}-C^{\beta}$	0.154	HNC^{α}	114°
$C^{\alpha}-H$	0.107	$XC^{\alpha}Y$	109.5°

diffraction studies and of the consideration of the tight packing of chains with a maximum number of hydrogen bonds. There are three such conformations. First, this is the *α-helix* shown in Fig. 4.2. It is characterized by rotation about the axis by 100° and displacement along the axis by 0.15 nm per peptide unit. Accordingly, there are 3.6 peptide units per each complete turn of the helix, the height of each turn being 0.54 nm. The helix is stabilized by hydrogen bond formation between the C=O group of one peptide bond and the H—N group of the preceding fourth peptide unit. Such bonds are formed between all the amino-acid residues, with the exception of proline (Pro) which contains no N—H group. The α-helix may be either right- or left-handed. In the first case, $\varphi = -132°$, $\psi = 123°$, and in the second, $\varphi = 228°$ and $\psi = 237°$ (the angles are measured from the planar *trans*-conformation of the chain). The conformation of the α-helix is determined, in particular, by the planar arrangement of the atoms in —CO—NH—.

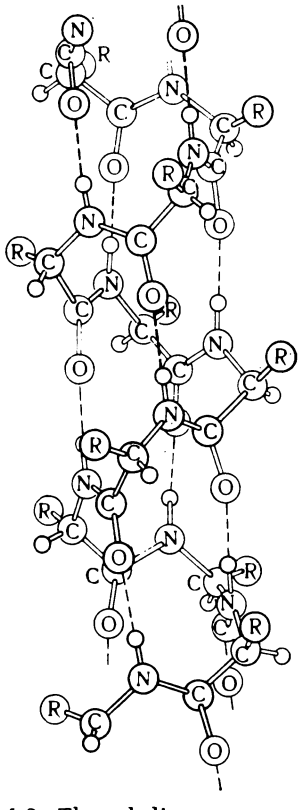


Fig. 4.2. The α-helix.

The second and third conformations with maximum saturation of the hydrogen bonds—the *parallel* and *antiparallel* β-*forms*—are shown in Figs. 4.3 and 4.4. These conformations are the conformations not of a single chain but of the assembly of chains that form a pleated-sheet structure. The chains in the β-forms have no planar, *trans*-structure (Fig. 4.5). In the parallel β-form $\varphi = 61°$ and $\psi = 239°$ and in the antiparallel β-form $\varphi = 380°$ and $\psi = 325°$.

Of importance is the possibility of a β-form being formed in a single polypeptide chain as a result of its regular twists. A schematic representation of such a *cross-β-form* is given in Fig. 4.6. The rotational angles at the bent points are different from those characteristic of the ordered portions.

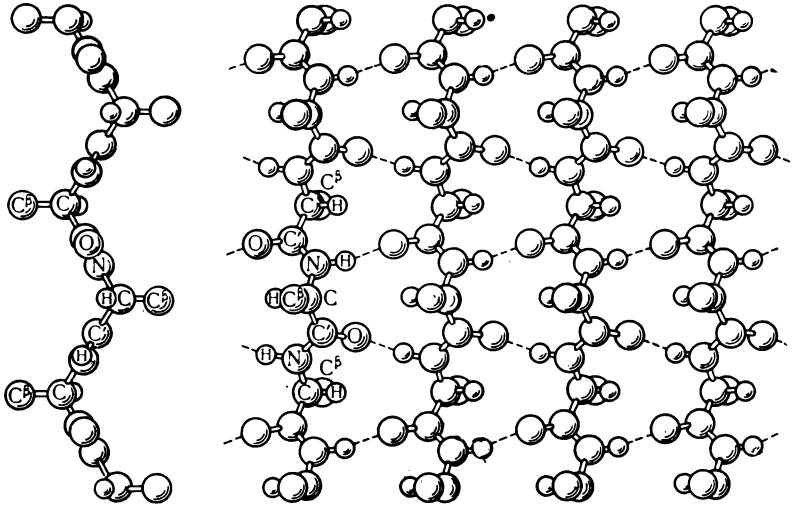


Fig. 4.3. The parallel β-form. Left—an individual chain.

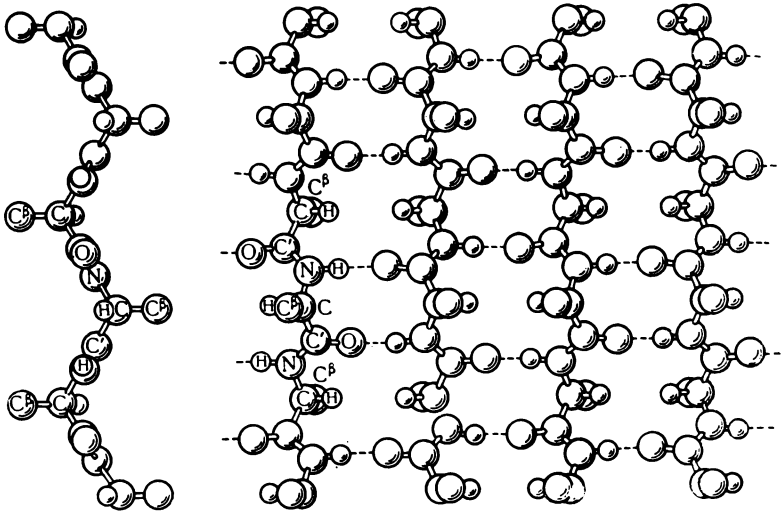


Fig. 4.4. The antiparallel β-form. Left—an individual chain.

Thus, the hydrogen bonds stabilize the fixed conformations of the polypeptide (protein) chain in solution. The fixed conformation is commonly referred to as the secondary structure of the chain. The presence of the secondary structure, which is periodic, means that the chain is similar to a crystal; the α-helix is similar to a one-dimensional and the β-form to a two-dimensional crystal.

The α- and β-forms are not the only ones. For example, fibrous proteins (Sec. 4.9) have other conformations.

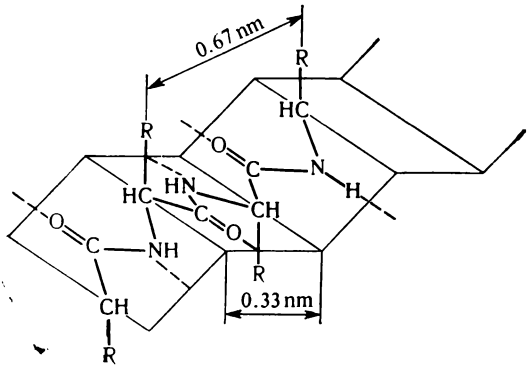


Fig. 4.5. The gross structure of the polypeptide chain in the β-form.

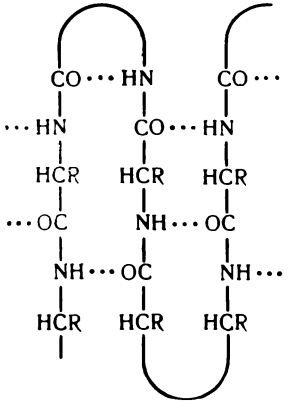


Fig. 4.6. Schematic representation of the *cross*-β-form.

Let us now consider the relationships between the energy of the polypeptide chain and the angles of internal rotation φ and ψ (so-called steric contour maps which resemble geodesic contour maps).

The conformational energy of the chain is determined by the weak interactions of nonbonded atoms. Because of the planar structure of the peptide group the rotational angles φ_i and ψ_i of the ith unit are practically independent of the angles φ_{i+1} and ψ_{i+1} of the adjacent unit. If the angles φ_i and ψ_i vary within the limits allowed by the overlap of the atoms of the peptide groups linked by the bonds of the ith and $(i + 1)$st units and if the angles φ_{i+1} and ψ_{i+1} vary simultaneously, then there is not such a combination of these four angles at which the steric interaction of the ith unit with the $(i + 2)$nd unit is possible. Therefore, the polypeptide chain has limited cooperativity, the short-range interactions in it being restricted to nearest-neighbour interactions. This allows one to deal separately with conformational energies for individual amino-acid

residues. The steric map for a particular residue depends substantially on the nature of its radical R. It may be presumed that the interactions in a given pair of peptide groups characterize the amino-acid residue that links these groups together. Ramachandran studied the dipeptide glycine-L-alanine and obtained a conformational (steric) map for alanine. The calculation was based on the assumption that atoms are rigid spheres having van der Waals radii determined from data on internuclear distances in molecular crystals. Table 4.2 gives internuclear distances most frequently observed in crystals and the minimum distances observed only in a few cases.

Table 4.2. The Contact Distances Between the Atoms in Polypeptides

Pair of atoms	Ordinary distance, nm	Minimum distance, nm	Pair of atoms	Ordinary distance, nm	Minimum distance, nm
C C	0.32	0.30	O N	0.27	0.26
C O	0.28	0.27	O H	0.24	0.22
C N	0.29	0.28	N N	0.27	0.26
C H	0.24	0.22	N H	0.24	0.22
O O	0.28	0.27	H H	0.20	0.19

Using the potential of rigid spheres, one can evidently obtain only the allowed and forbidden regions of values of φ and ψ. Figure 4.7 shows such a map for alanine. The solid lines delineate the regions allowed at ordinary internuclear distances, and the dotted lines those allowed at minimum distances. The map gives the values of φ and ψ for the right- and left-handed α-helices (α_D and α_L), the parallel and antiparallel β-forms (β_p and β_a) and for the fibrous protein collagen (C). Analogous maps have also been plotted for other residues. Naturally, as the size of the radical R increases the region of allowed values of φ and ψ diminishes.

More rigorous calculations have been carried out with the aid of the following potential (cf. page 68):

$$U(\varphi, \psi) = \frac{1}{2} U_{\varphi}^{\circ} (1 - \cos 3\varphi) + \frac{1}{2} U_{\psi}^{\circ} (1 - \cos 3\psi) + \\ + \sum_{i,j} U(r_{ij}) + U_C \qquad (4.1)$$

The potential $U(r_{ij})$ may be taken in the form of the Lennard-Jones (12-6) potential or the 6 — exp potential (page 68). The quantity U_C characterizes the Coulomb, electrostatic interaction between neighbouring pairs of peptide groups which is determined by their large dipole moments amounting up to 3.7 D.

The energy barriers U_{φ}° and U_{ψ}° are found from data obtained in the investigation of small model molecules. Evidently, $2.1 < U_{\varphi}^{\circ} < 5.5$ kJ/mole (2.1 is the energy barrier for $H_3C—CO—OH$, 5.5 for $H_3C—CO—Cl$) and $U_{\psi}^{\circ} < 6.3$ kJ/mole. These barriers are small, much lower than those for hydrocarbons.

The parameters for the Lennard-Jones (12-6) potential are listed in Table 3.2 (page 69). The calculation of U_C is complicated by the

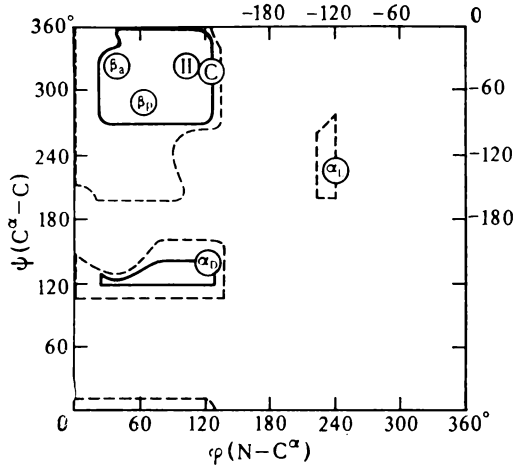


Fig. 4.7. A steric chart for L-alanine (Ramachandran).

fact that the value of dielectric constant ε that appears in Coulomb's law is unknown:

$$U_C = \sum_{i,\ j} \frac{e_i e_j}{\varepsilon r_{ij}} \tag{4.2}$$

where e_i and e_j are the charges on atoms i and j. If there are no other atoms or molecules of the solvent (water) between the interacting atoms, then $\varepsilon \approx 3$. This value corresponds to the hydrocarbon, organic environment.

For the atoms O and N linked by a hydrogen bond, use can be made of the function $U_{hb} = U_{ij} + U_C$, which is determined from the values of the hydrogen-bond energy and equilibrium O · · · H—N distances.

Expression (4.1) does not take account of the distortions of bond lengths and valence angles which are presumably small. This expression is valid for an isolated polypeptide chain. In the presence of an

aqueous environment, account should be taken of the change in free energy of the water molecules removed from the solvent shell when the atoms are in contact with one another. Strictly speaking, it is necessary to minimize the sum of the intramolecular potential energy and the free energy of the solvent.

Figure 4.8 shows the steric map for alanine. The numbers on the curves indicate the value (in kJ/mole) of the energy U [Eq. (4.1)].

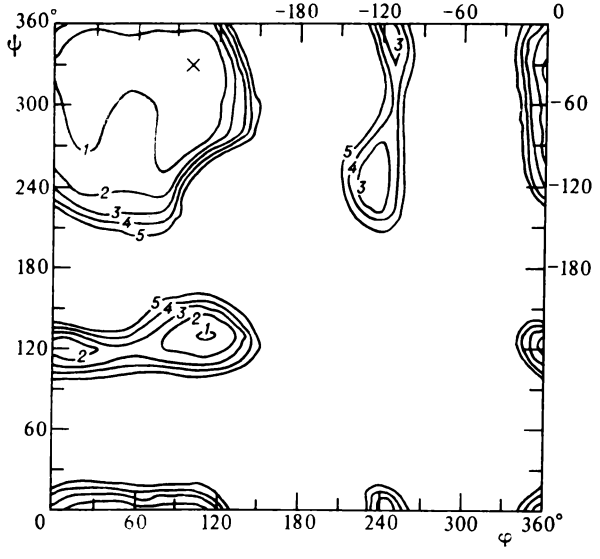


Fig. 4.8. A steric map for L-alanine (Flory).

The contour lines are drawn, which correspond to energies from 4 to 20 kJ/mole. The cross stands for the energy minimum. It will be instructive to compare the maps of Figs. 4.7 and 4.8.

4.3. The Hydrogen Bond and the Structure of Water

Thus, the secondary structures of protein chains are stabilized by hydrogen bonding. Hydrogen bonds also play a key role in the conformational structure of nucleic acids and carbohydrates.

Biopolymers function in an aqueous environment. The special properties of water, which make it an indispensable component of cells and organisms, are also determined by hydrogen bonds. Let us consider the nature and properties of the hydrogen bond (see page 62).

The presence of hydrogen bonds exerts a strong effect on the physical and physico-chemical properties of substances. The intermolecular hydrogen bonds determine the association of molecules. Associated substances are characterized by relatively high heats of vaporization, high melting and boiling points and by large differences in melting and boiling points. Let us compare four substances composed of *isoelectronic molecules* (Table 4.3). Methane is an unassociated substance, the other three are associated through hydrogen bonds. We shall compare the properties of two isomeric substances: ethanol and dimethyl ether (Table 4.4). Ethanol contains an OH group which forms a hydrogen bond. No such bonds are present in dimethyl ether.

Table 4.3. The Properties of Isoelectronic Substances

Substance	T_m, K	T_b, K	Heat of vaporization, kJ/mole	Molar volume, cm^3/mole
Hydrogen fluoride, HF	181	292	30.2	20.2
Water, H_2O	273	373	40.8	18.0
Ammonia, H_3N	195	240	23.4	20.8
Methane, H_4C	89	112	9.3	34.0

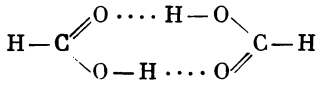
Table 4.4. The Properties of Ethanol and Dimethyl Ether

Substance	T_m, K	T_b, K	Heat of vaporization, kJ/mole
Ethanol, C_2H_5OH	161	351	42.8
Dimethyl ether, $(CH_3)_2O$	135	249	18.7

Substances associated by hydrogen bonds have large values of dielectric constants ε. For example, at 20°C the dielectric constant of H_2O is 80, that of HCN is 95, formamide, $HCONH_2$, 84, and of NH_3, 15.5. The dielectric constant of diethyl ether, $(C_2H_5)_2O$ is as low as 4.3.

The structural studies of crystals containing hydrogen bonds show that when two electronegative atoms A and B (O, N, F, Cl) are bound by a hydrogen atom, the distance between them, r_{AB}, is smaller than the sum of the van der Waals radii. The sums of these radii are as follows: 0.28 nm for two oxygen atoms, 0.30 nm for O and N, 0.33 nm for two nitrogen atoms. In the presence of hydrogen atoms the distance is reduce to: 0.255 nm in O—H · · · O; 0.280 nm in O—H · · · N, 0.288 nm in N—H · · · O, and 0.310 nm in N—H · · · N.

The hydrogen bond reveals itself distinctly in optical spectra and in NMR spectra. The characteristic vibration frequencies of groups containing a hydrogen atom, say O—H and N—H groups, decrease if hydrogen forms a hydrogen bond. For example, in monomeric formic acid, H—CO—OH, the vibrational frequency of the O—H group is 3682 cm^{-1}, and in the dimer stabilized by hydrogen bonds



it is equal to 3080 cm^{-1}. The other frequencies of the monomer are also changed. The infrared absorption bands of O—H groups are strongly broadened when a hydrogen bond is formed, and their intensities are enhanced. Hydrogen bonding affects electronic spectra, shifting $n\pi^*$-transitions (cf. page 155) to the short-wave side. In proton-magnetic-resonance (PMR) spectra there are observed chemical shifts caused by a hydrogen bond. The PMR spectroscopy is the most efficient method of studying hydrogen bonds.

The hydrogen bond energy is determined from the thermodynamic properties of substances, from their spectra, etc. The thermodynamic functions are expressed in terms of the equilibrium constant:

$$K = \frac{\text{Activity of substance}}{\text{Product of activities of reagents}} \approx \frac{[\text{A}-\text{H} \cdots \text{B}]}{[\text{AH}][\text{B}]} \qquad (4.3)$$

The square brackets signify concentrations. The free energy of formation of a hydrogen bond is given by

$$\Delta G = \Delta H - T\Delta S = -RT \ln K \qquad (4.4)$$

where ΔH is the enthalpy and ΔS is the entropy of bond formation. From Eq. (4.4) it follows that

$$\Delta H = RT^2 \left(\frac{\partial \ln K}{\partial T}\right)_p \qquad (4.5)$$

The values of ΔH are of the order of 12-30 kJ/mole: 11.8 for water, 25.6 for ice, 15.5-18.5 for ammonia, and 28.1-29.4 kJ/mole for HF.

A hydrogen bond is formed at carbon only in a few cases (HCN, chloroform, $CHCl_3$, mixed with pyridine, C_5H_5N, etc.),

Hydrogen bonding invariably links together two electronegative atoms. The hydrogen atom is not shared completely by the atoms A and B: the designation A—H · · · B, in which the solid line denotes a covalent bond and the dotted line a hydrogen bond, has a real meaning. The X-ray diffraction method does not allow one to directly determine the position of the hydrogen atom (see page 144), but it can be established by means of neutron diffraction (page 151). In the system O—H · · · O, for example, the covalent bond is much shorter and, hence, stronger than the hydrogen bond. The

following transitions occur in the system:

$$O-H \cdots O \rightleftharpoons O \cdots H-O$$

The curve showing the dependence of the energy on the position of the hydrogen atom in the system O—H · · · O (say, in the dimer of formic acid) has two symmetric minima separated by a barrier. Quantum mechanics permits us to consider the hydrogen bond from a quantitative viewpoint and to calculate its energy and length in reasonable agreement with experiment. The hydrogen bond is complex, its energy is made up of the contributions from electrostatic, dispersion and quantum-mechanical interactions (quantum-mechanical interactions are determined by the delocalization of the proton and electrons).

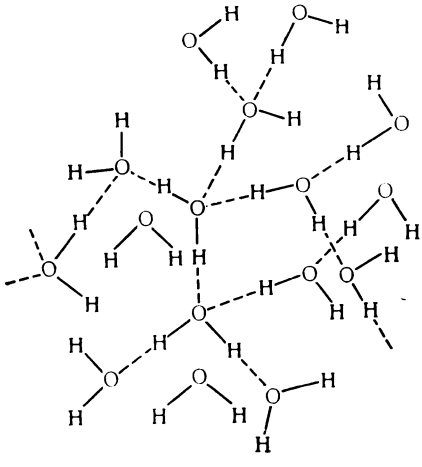


Fig. 4.9. The structure of water.

The structure of water and ice is determined by hydrogen bonds. A water molecule can form 4 hydrogen bonds with neighbouring molecules (Fig. 4.9), with a tetrahedral coordination. Ordinary ice (ice I) has a hexagonal crystalline lattice, each oxygen atom in the lattice residing in the centre of the tetrahedron, at the apexes of which there are located the neighbouring oxygen atoms. The O · · · O distances are equal to 0.276 nm. The elementary unit cell has four molecules. The molecular lattice of ice is very loose, with large open channels, since its *coordination number* is very small. This explains why ice is less dense than water. This property of ice is not unique; it is also exhibited by crystals of diamond, silicon, and germanium which have a similar structure.

The ice-like structure is retained also in liquid water, the number of broken crystalline portions being the greater, the higher the temperature. Speaking of the structure of water, one should take into account the time scale in which this structure is recorded. In the ice crystal the water molecules experience vibrations, rotations, and occasional translational displacements. A snapshot taken with an exposure time τ much shorter than the period of vibration, $\tau_{vib} \approx 2 \times 10^{-13}$ s displays an instantaneous or M-structure shown in Fig. 4.10*a*. During the time τ, much longer than the time τ_{vib}

but considerably shorter than the time of rotational diffusion, $\tau_{dif} \sim 10^{-5}$ s, the vibrations are averaged and the photograph will show regularly arranged but randomly oriented molecules—the so-called K-structure (Fig. 4.10*b*). And, finally, at $\tau \gg \tau_{dif}$, i.e., in an ordinary experiment, there will be obtained a diffusion-averaged K-structure (the D-structure) (Fig. 4.10*c*). In liquid water the M- and K-structures are similar to crystalline structures, but the D-structure is smeared by displacements of the molecules. If we place men-

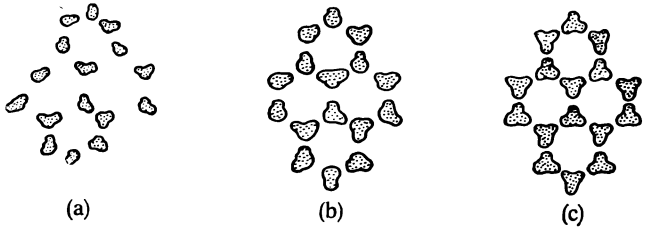


Fig. 4.10. The instantaneous (*a*), vibration-averaged (*b*) and diffusion-averaged (*c*) structures of water.

tally the camera over the water molecule and record the surrounding molecules during motion, we shall obtain the D-structure of the liquid which is its averaged K-structure.

Using various methods of investigation with different durations of experiments, τ, we can obtain information on the various types of structure (Fig. 4.11). The thermodynamic properties naturally characterize the D-structure of the liquid.

We note that the longest relaxation times in water are of the order of 10^{-5} s (τ_{dif} for ice). The occasionally published statements that water "remembers", as it were, what has happened to it—whether it was heated or frozen—are based on poorly conducted experiments. There is no "thermally activated" water and snow melt is devoid of any special properties.

The X-ray diffraction analysis of water gives a radial distribution function, i.e., a relative content of molecules which are at the various distances from one another. At temperatures from 4° to 120°C the principal peak of this function is gradually shifted from 0.282 to 0.294 nm. The coordination number in this interval is, on the average, 4.4 (for comparison, in liquid Ne the coordination number is 8.6 and in Ar it is equal to 10.5). Thus, water is quasi-crystalline and each molecule has, on the average, four neighbours. A number of theoretical models of the water structure have been proposed, which are consistent with these data; for this reason it is

so far difficult to decide on any one structural picture of liquid water. But we can consider the interpretation of the most important physical properties of water and, in the first place, the dependence of specific volume on temperature to have been established. The minimum of specific volume at 4°C is accounted for by the competition between two processes. The first process is the collapse of the ice open structure with a small coordination number (4), causing a decrease in volume. This is a continuation of the melting process. In one model the decrease in volume is associated with the filling

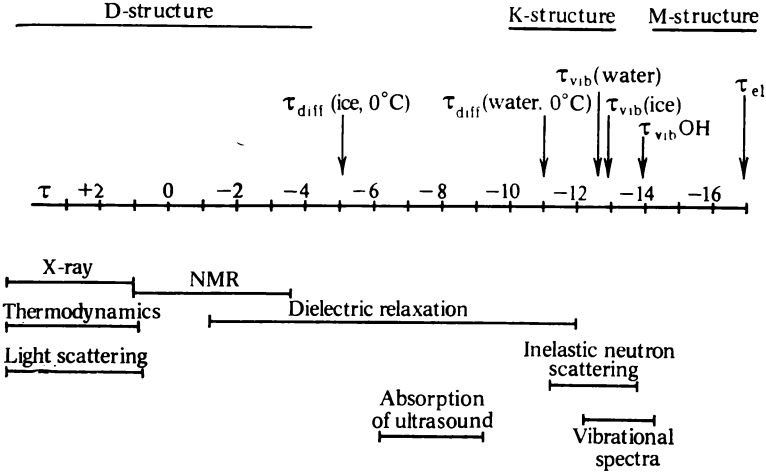


Fig. 4.11. The time intervals (log τ, τ in seconds) corresponding to various methods of investigation of the structure of ice and water.

of the empty space in the loose lattice by monomeric water molecules, while in another model the bending of hydrogen bonds leads to the approach of neighbouring molecules, i.e., to a decrease in volume. The second process, which predominates up to $T > 4$°C, is the thermal expansion of the liquid due to an increase in the amplitudes of anharmonic intermolecular vibrations.

4.4. Helix-Coil Transitions

The conformations of polypeptide chains stabilized by hydrogen bonds are stable only under certain conditions. A change of temperature, solvent, pH of the medium leads to order-disorder transitions, to the transformation of the regular conformation of the chain to

a random coil. These processes can be conveniently studied on model homopolymers—synthetic polyamino acids.

Many polyamino acids, say, polyglutamic acid (PGA) and its derivative poly-γ-benzyl glutamate (PBG)

$$\left[\begin{array}{l} -CO-CH-NH- \\ \quad\quad | \\ \quad\quad CH_2-CH_2-CO-O-CH_2-C_6H_5 \end{array}\right]_n$$

are present in solution in the form of α-helices, which is proved by their hydrodynamic and optical properties. Doty has found that

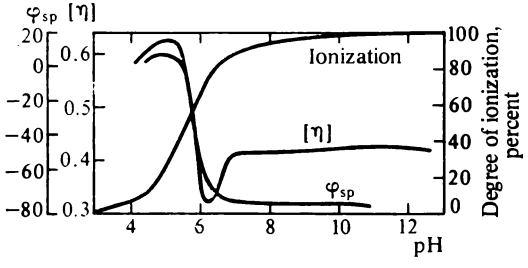


Fig. 4.12. The dependence of the degree of ionization, [η] and φ_{sp} on the pH of the medium in the region of the helix-coil transition in polyglutamic acid (PGA).

helix-coil transitions are very sharp and similar to phase transitions—there occurs a sort of melting of α-helices of one-dimensional crystals. Figure 4.12 shows the dependence of the degree of ionization, the intrinsic viscosity [η] and specific optical rotation φ_{sp} of PGA on the pH of the medium. Near pH = 6 the values of [η] and φ_{sp} fall off sharply and the degree of ionization increases. At pH < 6 polyglutamic acid has a helical conformation and at pH > 6, a coiled conformation. In contrast to PGA, poly-γ-benzyl glutamate (PBG) dissolves in organic solvents. When dissolved in dichloroethane, chloroform, and formamide PBG forms rigid rods—α-helices. On the contrary, in solvents whose molecules form hydrogen bonds with PBG (in trichloroacetic acid, etc.), PBG is present in a coiled form. As the composition of the solvent (say, the binary mixture $CHCl_3 + CH_2ClCOOH$) is changed gradually within a certain narrow region (at 80 percent of $CHCl_3$) the helix is converted to a coil.

The sharpness of the transition and the sigmoidal character of the corresponding curves (for example, the curve of φ_{sp} in Fig. 4.12) are an indication of the cooperative nature of the transition. The factor responsible for cooperativity is evident from the considera-

tion of the α-helix structure. The conformations of peptide units are interdependent, since the hydrogen bond between the C=O group of the ith unit and the N—H group of the $(i-4)$th unit fixes the conformations of the $(i-1)$st, $(i-2)$nd, and $(i-3)$rd units. In order to liberate a peptide unit, which will give a gain in entropy, it is necessary to break at least three hydrogen bonds, one after another, which requires the expenditure of energy. To reduce the free energy G, many units must be liberated, i.e., *cooperativity* is required.

The thermodynamic condition for the melting of a crystal consists in the equality of the free energies of the crystal and melt—an α-helix and a coil in our case. We have

$$G_\alpha = H_\alpha - T_m S_\alpha = H_{\text{coil}} - T_m S_{\text{coil}} = G_{\text{coil}} \tag{4.6}$$

or

$$\Delta G = G_{\text{coil}} - G_\alpha = \Delta H - T_m \Delta S = 0 \tag{4.7}$$

The change in enthalpy is compensated for by the entropy change. From Eq. (4.7) it follows that

$$T_m = \frac{\Delta H}{\Delta S} \tag{4.7a}$$

The statistical theory (to be considered at a later time) shows that the fraction of helical units of such a cooperative system is given by the following formula:

$$\theta = \frac{s^n}{1+s^n} \tag{4.8}$$

where $1 < n < N$ (N is the number of units in the α-helix) and

$$s = \exp\left(-\Delta G/RT\right) \tag{4.9}$$

has the meaning of the equilibrium constant for the formation of a hydrogen bond in the unit that follows the hydrogen-bonded unit. Formula (4.8) describes the cooperative melting of the α-helix if $n \gg 1$. In this case at $s = 1$ a sharp transition occurs: $\theta \approx 0$ at $s < 1$; $\theta = 0.5$ at $s = 1$ (i.e., at $\Delta G = 0$) and $\theta = 1$ at $s > 1$. Thus, formula (4.8) is in accord with the thermodynamic melting condition (4.7), i.e., with melting at a certain definite temperature.

But under what conditions is the value of n much greater than unity? The cooperativity is high if the free energy required for a single break to occur in the sequence of hydrogen bonds is high, i.e., if the equilibrium constant σ for such a process is small:

$$\sigma = \exp\left(-G_{\text{br}}/RT\right) \tag{4.10}$$

Calculations show that

$$n \approx \frac{1}{\sqrt{\sigma}} \tag{4.11}$$

The transition is the sharper the smaller is the cooperativity parameter σ. The cooperativity is at a maximum if $G_{br} \to \infty$ and $\sigma \to 0$. In such a case $n \to N$. On the contrary, no cooperativity is present if G_{br} tends to zero and $\sigma \to 1$. In this case $n \to 1$ and, instead of formula (4.8), we get

$$\theta = \frac{s}{1+s} \qquad (4.12)$$

The quantity θ changes smoothly, without kinks—no melting occurs. Figure 4.13 shows the theoretical curves of the function $\theta(s)$ at various values of σ.

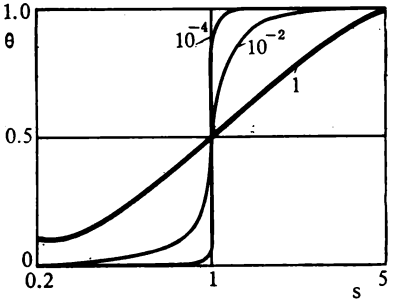


Fig. 4.13. Theoretical curves of θ versus s at different values of σ.

We shall give a statistical-mechanical derivation of these relations based on the Ising model and the matrix method (page 79). Each unit of the polypeptide chain may be in a non-bonded state (the symbol $\mu_i = 0$) and in a state bound by hydrogen bonds ($\mu_i = 1$). The free energy of the chain depends on a set of μ_i values, the conformations of four successive units being interdependent. Therefore, the free energy of the chain is given by

$$G\{\mu_i\} \equiv G(\mu_1, \mu_2, \ldots, \mu_N) = \sum_{i=1}^{N} \tilde{G}_i(\mu_{i-3}, \mu_{i-2}, \mu_{i-1}, \mu_i) \qquad (4.13)$$

As before (page 80), the chain is assumed to be long, $N \gg 1$ and the end effects are neglected. It is obvious that the free energy required for the release of one or two units enclosed between bonded units must be very high, since there is actually no release in this case, the units remaining in the helix. The enthalpy is expended without any gain in entropy. It may be assumed that

$$\tilde{G}(\mu_{i-3}, 1, 0, 1) \to \infty, \quad \tilde{G}(1, 0, 0, 1) \to \infty \qquad (4.14)$$

These states practically do not participate in the partition function, which is given by

$$Z = \sum_{\{\mu_i\}} \exp(-G\{\mu_i\}/RT) \qquad (4.15)$$

Distracting from these states, we may limit ourselves to the consideration of only two neighbouring units and simplify expression (4.13):

$$G\{\mu_i\} = G(\mu_1, \mu_2, \ldots, \mu_N) = \sum_{i=1}^{N} \tilde{G}(\mu_{i-1}, \mu_i) \qquad (4.16)$$

Here we are dealing with only four values of $\tilde{G}$: $\tilde{G}(0,0)$, $\tilde{G}(0, 1)$, $\tilde{G}(1, 0)$, and $\tilde{G}(1, 1)$. We assume that the free energy of a unit in a free state ($\mu_i = 0$) is equal to zero—the unit has melted. We have

$$G_{\text{free}} = \tilde{G}(0, 0) = \tilde{G}(1, 0) = 0 \tag{4.17}$$

For a unit in a bound state ($\mu_i = 1$)

$$G_{\text{bound}} = \tilde{G}(1, 1) = \Delta G \tag{4.18}$$

And, finally, for the state $\mu_{i-1} = 0$, $\mu_i = 1$

$$\tilde{G}(0, 1) = G_{\text{bound}} + G_{\text{br}} \tag{4.19}$$

If we assume that three rotational isomers can be formed through the rotation of a free unit about each of the bonds, then

$$G_{\text{br}} = 4RT \ln 3 \approx 10.5 \text{ kJ/mole} \tag{4.20}$$

These quantities are incorporated into Z as exponents: G_{free} gives the factor 1, G_{bound} the factor s [formula (4.9)], and G_{br} gives the factor σ [formula (4.10)]. So we have (cf. page 80)

$$Z = \sum_{\{\mu_i\}} \prod_{i=1}^{N} s^{\mu_i} \sigma^{\mu_i(1-\mu_{i-1})} = \text{Sp}\,(\boldsymbol{P}^N) = \lambda_1^N + \lambda_2^N \tag{4.21}$$

The matrix $\boldsymbol{P}$ has the following form:

$$\boldsymbol{P} = \begin{array}{c|cc} \mu_{i-1} \backslash \mu_i & 0 & 1 \\ \hline 0 & 1 & \sigma s \\ 1 & 1 & s \end{array} \tag{4.22}$$

Its characteristic equation is

$$(1 - \lambda)(s - \lambda) = \sigma s \tag{4.23}$$

The eigenvalues of the matrix are

$$\lambda_{1,2} = \frac{1}{2}(1+s) \pm \left[\frac{1}{4}(1+s)^2 + s\sigma\right]^{1/2} \tag{4.24}$$

With maximum cooperativity $\sigma = 0$ and $\lambda_1 = 1$, $\lambda_2 = s$. We have

$$Z = 1 + s^N \tag{4.25}$$

and

$$\theta = \frac{1}{N}\frac{\partial \ln Z}{\partial \ln s} = \frac{s^N}{1+s^N} \tag{4.26}$$

If there is no cooperativity, then $G_{br} = 0$ and $\sigma = 1$. In such a case $\lambda_1 = 1 + s$, $\lambda_2 = 0$, and we obtain formula (4.12). In intermediate cases, where $0 < \sigma < 1$, we get formula (4.8). We introduce the notation $s^n = \tilde{s}$. This quantity has the meaning of the equilibrium constant for a unimolecular reaction which involves n units, i.e., the effective enthalpy change:

$$\Delta H_{\text{eff}} = n\Delta h \tag{4.27}$$

where Δh is the enthalpy change when one unit is liberated. At $N \gg 1$ we may restrict ourselves to the larger root of (4.24) in expression (4.21):

$$Z \approx \lambda_1^N \tag{4.28}$$

So we have

$$\theta = \frac{1}{N}\frac{\partial \ln Z}{\partial \ln s} \approx \frac{\partial \ln \lambda_1}{\partial \ln s} = \frac{\lambda_1 - 1}{2\lambda_1 - 1 - s} \tag{4.29}$$

Equating this expression to expression (4.8), with s^n being given by

$$s^n = \tilde{s} = \exp\left(-\frac{\Delta H_{\text{eff}} - T\Delta S_{\text{eff}}}{RT}\right) \tag{4.30}$$

we find

$$\tilde{s} = \frac{\lambda_1 - 1}{\lambda_1 - s} \tag{4.31}$$

and

$$\Delta H_{\text{eff}} = -\frac{d \ln \tilde{s}}{d(1/RT)} = -\frac{d \ln s}{d(1/RT)}\frac{d \ln \tilde{s}}{d \ln s} = \Delta h \frac{d \ln \tilde{s}}{d \ln s} \tag{4.32}$$

Calculation gives

$$\frac{d \ln \tilde{s}}{d \ln s} \approx \frac{s}{(\lambda_1 - 1)(\lambda_1 - s)} \frac{2(\lambda_1 - 1)(\lambda_1 - s) + \sigma(1 - s)}{2\lambda_1 - 1 - s} \tag{4.33}$$

and in the transition region, where $s \approx 1$

$$\frac{d \ln \tilde{s}}{d \ln s} \approx \frac{2s}{2\lambda_1 - 1 - s} \approx \frac{1}{\sqrt{\sigma}} \tag{4.34}$$

Hence,

$$\Delta H_{\text{eff}} \approx \frac{\Delta h}{\sqrt{\sigma}} \tag{4.35}$$

that is,

$$n \approx \frac{1}{\sqrt{\sigma}} \tag{4.36}$$

The quantity ΔH_{eff} can be found from the slope of the curve of θ versus $1/RT$ in the transition region, where $\tilde{s} \approx 1$

$$\frac{d\theta}{d(1/RT)} = -\frac{d\theta}{d\ln \tilde{s}} \Delta H_{\text{eff}} = -\frac{\tilde{s}}{(1+\tilde{s})^2} \Delta H_{\text{eff}} \approx -\frac{1}{4} \Delta H_{\text{eff}} \quad (4.37)$$

Optical methods enable one to determine θ (T) and, hence, $\Delta h/\sqrt{\sigma}$. To find the values of Δh and σ separately, use may be made of the dependence of T_m on the degree of polymerization N.

It has been found experimentally that $\sigma = 2 \times 10^{-4}$ and the enthalpy change upon liberation of one unit, Δh, is about 4 kJ/mole for PBG. The same value of σ is characteristic of PGA and poly-L-lysine. The value of Δh, however, is in these two cases negative and low, about -0.29 kJ/mole. If $\Delta h > 0$, the helical state is less favourable than the coiled state and the helix-coil transition occurs with fall and not with rise of temperature.

The theory of transitions brought about by a change of the solvent or of the pH of the medium is worked out on the same grounds. In the first case it is necessary to take into account that each unit may be not in two but in three states: in a free state, in a state bound by an intramolecular hydrogen bond and in a state bound by a hydrogen bond with the solvent molecule. The dependence of the transition on pH is determined by the polyelectrolyte nature of polypeptides such as PGA. Each link can be in one of the four states, since it may be either free or bound, either charged or uncharged. For polyampholytes the shape of the θ (pH) curve depends on the sequence of the cationic and anionic units. Theoretical calculations of the transitions are in agreement with experiment.

The β-form-coil transitions are of a somewhat different nature, since the β-form is two dimensional. The theory of such transitions has been developed by Birshtein.

Thus, the weak interactions in the model polypeptide systems lead to their sharply pronounced cooperative behaviour, to the occurrence of transitions which are similar to phase transitions.

4.5. The Protein Globule and Hydrophobic Interactions

In contrast to a monotonous polyamino acid, a protein contains various residues, including Pro, which are not capable of forming hydrogen bonds. The secondary structures—α-helices and β-forms—are present in the protein only partially, they alternate with disordered portions, in which the protein chain possesses a considerable flexibility. As a result, the protein macromolecule is coiled into a globule and assumes a definite three-dimensional tertiary structure. It is this structure that is biologically functional. (The fibrous proteins will be described in Sec. 4.9.)

In a number of cases, the protein has a quaternary structure—the protein molecule or the supermolecular protein system is composed of a certain number of globules. Examples are as follows: the hemoglobin molecule contains 4 globules of two species; the protein coat of the tobacco mosaic virus consists of 2000 identical globules. The protein is a multilevel system.

The nature of the structure at each of the levels of organization is determined by the geometric properties of the structures of the preceding level, the forces of interaction of their elements and by the interaction with the environment. The appearance of a "higher" structure occurs, as it were, automatically as a result of the *self-assembly* of the system. We shall deal with the self-assembly of the globule in Sec. 4.7.

The solution of the problems of self-assembly would mean, for example, the possibility of prediction of the macroscopic structure of the muscle on the basis of data on the chemical structure of its proteins. The explanation of the structure of molecular crystals based on the knowledge of the structure of the molecules of which these crystals are composed solves the problems of self-assembly in a much simpler case.

We have already mentioned the difference between a polymer globule and a random coil (page 84). In the protein, because of the polyfunctionality of amino-acid residues, various forces take part in the formation of a globule. The diversity of the units and of interactions involved determines the structure of the globule—the "aperiodicity" of the crystal.

The sole strong interactions in the protein globule are chemical disulphide linkages Cys—S—S—Cys. The presence of several disulphide "cross-links" between the units of one or more chains (say, insulin) imposes limitations on the possible conformations. However, one should not speak of the globule if these interactions were reduced to disulphide bridges. In this case, the protein chain would be similar to the chain in natural rubber vulcanized with sulphur. Natural rubber retains the properties of a random coil. The globule is formed by weak forces—electrostatic and van der Waals forces, hydrogen bonds, and, in the first place, *hydrophobic interactions.*

Globular proteins function in a water environment, which determines to a considerable extent the higher levels of their structure. The effect of water must be taken into account in the physics of the protein.

Water affects the hydrogen bonds. The gain in free energy upon formation of intramolecular hydrogen bonds is, however, insignificant as compared to hydrogen bonds with water molecules. The role of water in the stabilization of the globule is of a different kind.

The hydrophobic effect is the only organizing force based on the repulsion of the solvent and not on the mutual attraction of the

elements of the system. Water repels nonpolar molecules, particularly hydrocarbons—kerosene or an oil does not mix with water.

In an equilibrium solution or mixture the chemical potentials of the components are equal. In the case of any two environments, a and b, for molecules of a given substance

$$\mu_{(a)} = \mu_{(b)} \tag{4.38}$$

or

$$\mu^0_{(a)} + RT \ln x_{(a)} = \mu^0_{(b)} + RT \ln x_{(b)} \tag{4.39}$$

where $x_{(a)}$ and $x_{(b)}$ are the mole fractions of the substance in media a and b, which are in equilibrium; $RT \ln x$ characterizes the entropy of mixing; μ^0 is the standard chemical potential which incorporates the free energy of all molecular motions of an isolated substance and the energy of specific interactions with surrounding molecules. Imagine a system consisting of a nonpolar liquid, say CCl_4, and water. Carbon tetrachloride is immiscible with water and the system undergoes phase separation. We introduce a nonpolar substance (say, a hydrocarbon) into the system. This substance dissolves preferentially in a nonpolar solvent. At equilibrium, from Eq. (4.39) it follows for the hydrocarbon that

$$\mu^0_{CCl_4} - \mu^0_{H_2O} = RT \ln \frac{x_{H_2O}}{x_{CCl_4}} < 0 \tag{4.40}$$

since $x_{CCl_4} \gg x_{H_2O}$. Indeed, for hydrocarbons the differences in standard chemical potentials (4.40) are negative and are about 3570 J/mole for CH_2 groups and 8400 J/mole for CH_3 groups. The hydrophobic effect is determined by the properties of water—hydrophobic substances are indifferent towards nonpolar organic solvents.

Molecules that contain both polar and nonpolar groups arrange themselves in such a fashion that the former groups contact with water and the latter move away from the water environment. Langmuir has shown that it is in this way that the monomolecular layers of fatty acids are built on the water surface—the polar carboxyl groups of the molecules are immersed in water, while the nonpolar hydrocarbon radicals project outward towards the surface. The same situation determines the structure of micelles in aqueous colloidal solutions of soaps: the hydrophobic groups arrange themselves in the interior of the micelle and the hydrophilic groups are directed towards its surface.

The physical nature of hydrophobic interactions is specific. The poor solubility of hydrocarbons in water is associated not with the increase in enthalpy of the system but with the decrease in its entropy. Accordingly, the solubility of a hydrocarbon in water decreases rather than increases upon heating. The enthalpy decreases too, but in the overall balance of the free energy this effect is overlapped

by the entropy effect. For example, when butane, C_4H_{10}, is dissolved in water at 298 K, its entropy decreases by 96.6 J/(mole·K) and the enthalpy by 4200 J/mole. As a result, the free energy increases by

$$\Delta G = \Delta H - T\Delta S = -4200 + 298 \cdot 97 \approx 33{,}000 \text{ J/mole} \quad (4.41)$$

The dissolution of the hydrocarbon in water may be interpreted as the incorporation of nonpolar molecules into structured (ice-like) regions of water and into more compact unstructured regions. In the first case, the nonpolar molecules arrange themselves in the cavities of the loose structure, forming *clathrates*; additional contacts between the molecules and water lead to a decrease in energy. In the second case, an increase in the number of hydrocarbon-water contacts brings about a decrease in the number of water-water contacts, which corresponds to a lower energy. As a result, the enthalpy is increased. Hydrocarbons dissolve better in structured cross-linked regions and the equilibrium is shifted to the side of increased structurization (the formation of "icebergs"), which means a decrease in entropy. This is a qualitative picture. No hydrophobic theory has been developed as yet, but good theoretical models are available at present.

Among the amino-acid residues of a protein there are polar and nonpolar as well as hydrophobic and hydrophilic residues. Since the protein chain possesses some flexibility (due to the presence of disordered regions), it is capable of coiling into a globule, so that the hydrophobic residues are predominantly in contact with one another and not with water. In other words, the central region of the globule (its core, so to say) must be hydrophobic ("fatty"), and the outer surface hydrophilic ("soapy").

In fact, globular proteins undergo denaturation, i.e., pass into a disordered coil-like state, under the influence of weakly polar organic solvents, which form hydrogen bonds less easily than water. The action of such solvents depends on the contacts with nonpolar amino-acid residues of the protein and thereby on the disturbance of hydrophobic interactions. The denaturing effect of alcohols on proteins increases with increasing size of the aliphatic radical. The strong denaturing action of urea is also accounted for by a weakening of hydrophobic interactions.

It is exactly because hydrophobic residues are pushed into the globule that the water surrounding the protein does not undergo additional structurization. Of course, the water molecules interact with polar residues on the globule surface, but this interaction resembles ordinary solvation.

Fisher has given a crude estimate of the effect of hydrophobic residues on the shape of the globule. He classified all amino-acid residues into two groups: polar and hydrophilic (Arg, Asp, His, Glu, Lys, Ser, Tyr, Thr) and hydrophobic (the remaining 12 amino-

acid residues). Considering that all the residues are approximately of the same volume, one can arrive at the shape of the globule if one knows the percentage content of both types of residues. With the volume being specified the sphere has the smallest surface area. If the number of hydrophilic residues is enough to cover the surface of the spherical hydrophobic core, then the globule has a spherical shape. If this number is larger, the globule assumes the form of an ellipsoid. Finally, if the number of hydrophilic residues is insufficient and they cannot cover the core of the globule, there will remain unprotected hydrophobic regions. In this case the globules will stick together, forming a quaternary structure. These conceptions are in crude agreement with experiment.

However, the actual picture is, of course, more complicated. It is impossible to replace 20 types of residues, each with its individual properties, by only two types. One should speak in such a case of the degree of hydrophobicity of the residue and introduce a quantitative measure for it. Tanford has suggested the free energy change, ΔG, per each side group (the radical R) of a free amino acid upon transfer from ethanol into water. Table 4.5 presents the relative values of ΔG determined experimentally by Tanford; the value of ΔG for Gly is taken as zero since Gly contains no side group. The first 10 residues may provisionally be considered to be hydrophobic and the second 10 hydrophilic. This classification does not coincide with the classification according to polarity, i.e., dipole moments. The strongly polar Arg is as hydrophobic as the nonpolar Ala due to the presence of a large hydrocarbon residue.

Table 4.5. The Hydrophobic Nature of Amino-Acid Residues

Residue	ΔG, J/mole	Residue	ΔG, J/mole
1. Trp	12,600	11. Ala	3,070
2. Ile	12,500	12. Arg	3,070
3. Tyr	12,100	13. Cys	2,700
4. Phe	11,100	14. Glu	2,300
5. Pro	10,900	15. Asp	2,270
6. Leu	10,200	16. Thr	1,850
7. Val	7,100	17. Ser	170
8. Lys	6,300	18. Gly	0
9. His	5,900	19. Asn	-40
10. Met	5,500	20. Gln	-420

The degree of hydrophobicity of amino-acid residues provides information on the stabilization of the globule in a water environment. It is, however, necessary to take into account the fact that hydrophobic residues may also be present on the globule surface if

they are adjacent to hydrophilic residues in the chain. One must consider the actual structure of the globule and the entire balance of interactions taking place in it.

Hydrophobic interactions play a determinative role for the structure and properties of biological membranes and membrane proteins (see Chapter 10).

4.6. Relationship Between the Primary and the Three-Dimensional Structure of Proteins

The building of biological molecular and supermolecular structures invariably involves two stages: the *biosynthesis* of large and small molecules and the *self-assembly* of the structure. Both these stages are based on *molecular recognition*, which is realized due to weak interactions.

The problem of self-assembly is of primary importance in application to proteins. Biosynthesis, i.e., the formation of the primary structure of the protein (see Chapter 8), is coded genetically. Biologically functional, however, is the native spatial structure of the protein molecule which is formed as a result of self-assembly. Natural selection of proteins occurs at the level of the three-dimensional—tertiary and quaternary—structures. Molecular biology and molecular genetics would be devoid of meaning if there were not an unambiguous, or degenerate, relationship between the genetically predetermined primary structure of the protein and its spatial structure (see Sec. 7.1).

The existence of such a relationship is deduced directly from experiments on the *renaturation* of proteins (Anfinsen). In a number of cases it was possible to observe the recovery of the native structure and of the functionality of a denatured protein. Upon slow "annealing" of a protein coil there takes place the self-assembly of the organized globule.

The protein chain may have an enormous number of conformations. To find the unique conformation that corresponds to the absolute minimum of free energy by running over all possible conformations is impossible. This task is evidently circumvented by nature too since it would require a large period of time, while the self-assembly of a protein globule takes 1 s. The main idea of all present-day investigations devoted to the prediction of the globular structure—an aperiodic crystal—based on the knowledge of the primary structure of the chain consists in that the native globule is the final result of self-assembly which does not necessarily correspond to the global (absolute) minimum of free energy. To deduce the structure of the native globule, one must start from a certain hierarchy of structures. The protein can be separated into helical or extended structural segments joined together by various folds or loops. Two or three

neighbouring structural segments in the chain form elementary complexes: "hairpins" consisting of antiparallel α-helices, antiparallel β-hairpins, and parallel β-hairpins covered by an α-helix. This gives rise to a *domain*, i.e., a compact structure built up of several neighbouring elementary complexes and structural segments. The globules of small proteins are composed of a single domain, those of

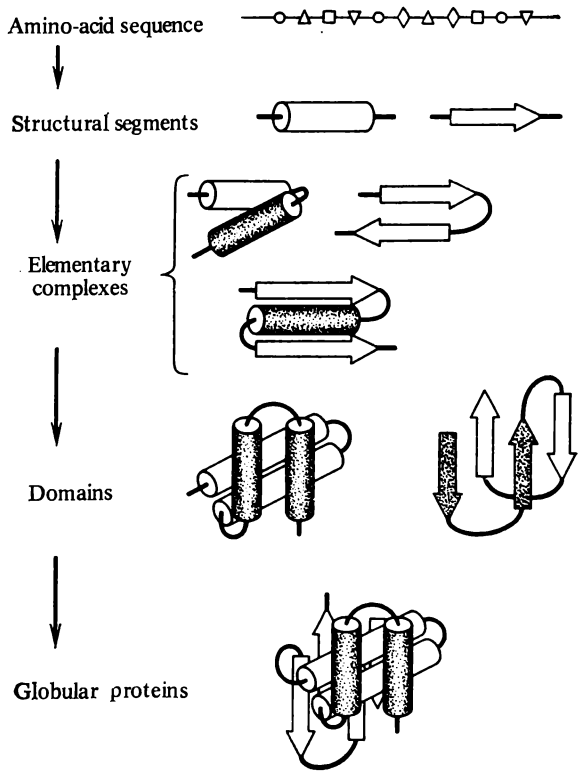


Fig. 4.14. The hierarchy of protein structures.

large proteins consist of several domains. This sequence of structure is shown schematically in Fig. 4.14. Thus, the twisting of a protein is supposed to proceed by a block mechanism—simpler structures of a lower level in the sequence serve as building blocks for the formation of higher structures (Ptitsyn).

Lim has proposed the stereochemical theory of the secondary (α- and β-) structure of globular proteins. This theory considers the

distribution of three types of residue over the surface of a structural portion, namely hydrophobic residues, residues with long polar side groups and residues with short polar groups. Only some of the ways of distribution of these residues over the surface of α-helices or of extended portions make it possible for them to be packed together tightly in the globule. Long hydrophilic side groups arranged at the boundary of the hydrophobic region and in the interior or on the edges of the helix shield the hydrophobic core of the globule against water. The fully hydrophobic portions invariably have either an α- or a β-form, while fully hydrophilic portions are disordered. It has been possible, on the basis of these considerations, to predict the degree of α-helicity of several proteins in satisfactory agreement with experiment.

The problem of self-assembly is not, however, reduced to geometry—it is a problem of physical dynamics. The secondary structure can serve as a block in self-assembly if, first, it is formed much faster than the tertiary structure; second, if it exists for a sufficiently long time and, third, if it is sufficiently large and hydrophobic to be involved in a strong hydrophobic interaction. Both the α-helices and β-forms meet these requirements. For a theoretical calculation of the secondary structure it is necessary that the parameters of equilibrium (the values of s; see page 110) between the various possible structures for all the residues and an adequate mathematical apparatus be available. Many, but not all, molecular parameters can be obtained by studying helix-coil and β-form-coil transitions in model polyamino acids. Ptitsyn and Finkelshtein have obtained the values of s_α and s_β for all amino-acid residues. Hydrophobic residues are characterized, as a rule, by the values of s_α and $s_\beta > 1$, i.e., they stabilize the α- and β-forms, whereas nearly all short hydrophilic and all charged residues (and also Gly and Pro) destabilize the α- and β-forms. The folds and loops of the chain are enriched with such destabilizing residues. Thus, the structure of the α- and/or β-portions in the extended chain is stabilized by interactions between hydrophobic groups that form a continuous hydrophobic surface. The distribution of hydrophobic groups that favours the formation of an α- or β-portion in the extended chain simultaneously provides their ability to be incorporated into a compact globule. There occurs the stabilization by local and then by long-range interactions.

The most probable secondary structures formed in a stretched-out protein chain predetermine the secondary structure of the globule.

The system formed is highly cooperative, and in the corresponding calculations one can make use of the apparatus of the *Ising model* (see pages 79 and 111).

The approach described above has enabled the three-dimensional structures of a number of proteins to be found. Figure 4.15 shows

a theoretical model of myoglobin. It is in good agreement with the structure established by the X-ray diffraction technique.

The self-assembly of the globule is characterized by two principal ways: the formation of a planar β-structure (a sheet) with an α-helix being subsequently stuck to it and the formation of a β-hairpin

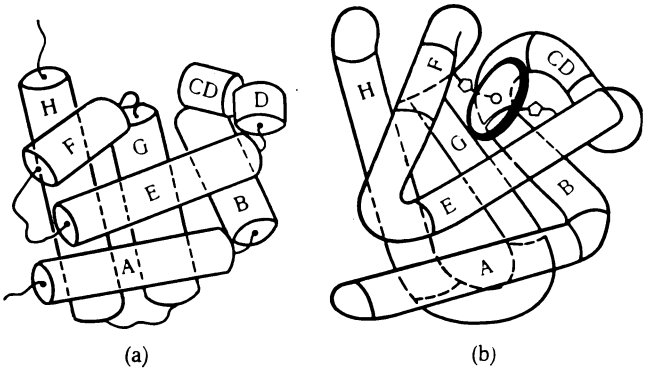


Fig. 4.15. A theoretical model of myoglobin (*a*) and its actual structure (*b*).

which is then folded in half. The entire self-assembly process involves three stages:

1. The formation of a fluctuating secondary structure in an extended protein chain which determines the type of the secondary structure of the native protein.

2. The initiation of the folding of the tertiary structure which is accomplished by an "initiating hairpin" from the neighbouring α- and/or β-portions of the chain. The formation of such a hairpin is the most unfavourable stage of the folding process. Therefore, there is only one "initiating hairpin" per each domain of the protein.

3. The growth of the tertiary structure which occurs through addition of neighbouring portions in the chain. Since the globule consists of layers formed by either α- or β-structural segments, and the disordered bridges between them pass along the boundaries of the layers, the neighbouring structural segments in the chain are antiparallel. The bridges do not intersect.

The one-centre mechanism of self-organization and the absence of intersections of the bridges sharply reduce the number of possible spatial structures. The prediction of the structure is simplified for purely α-helical and purely β-structural proteins. The initiation of an α-helix requires the consumption of free energy, but no free energy is needed for its growth. The formation of a single long helix

is therefore more favourable than the formation of several short ones. The longer is the helix, the more probable is the formation of a double-helical hairpin. In an analogous way, the formation of a long β-hairpin is more preferential than that of several short ones. The complex may bend or fold in chain portions situated between

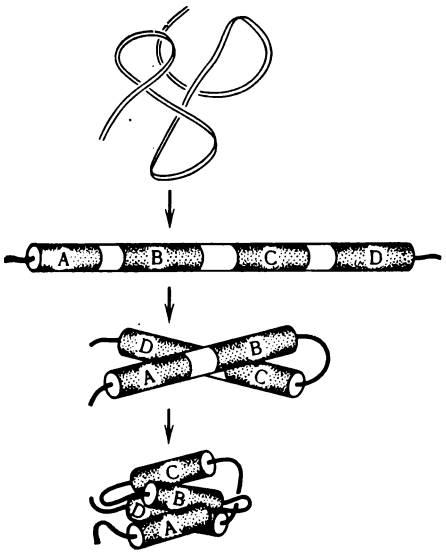


Fig. 4.16. Initiation of tertiary structure for an α-helical protein.

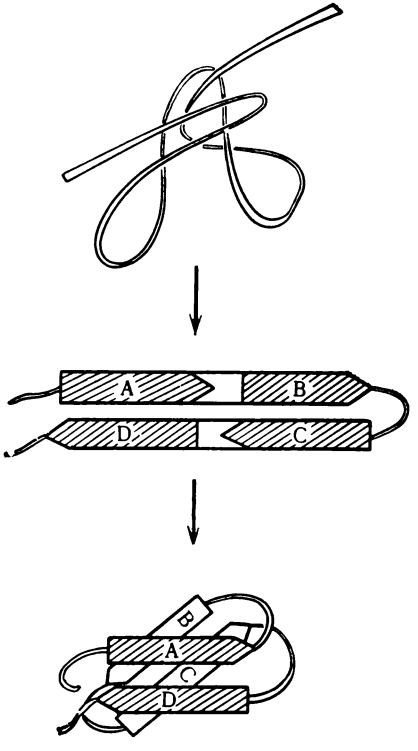


Fig. 4.17. Initiation of tertiary structure for the β-form.

the native structural segments. Figures 4.16 and 4.17 present the schemes of initiation of the tertiary structure for alpha-helices and beta-forms, respectively. The theory just described has furnished good agreement with experiment for a number of proteins, notably for ribonuclease. The theory agrees well with the cooperative nature of denaturation.

Despite the advances made in the works described above, the problem of the structure of the globule cannot yet be considered to have been solved. We have limited ourselves to the consideration of hydrophobic interactions. The electrostatic interactions between ionogenic polar side groups and water molecules are also impor-

tant. If the protein contains a cofactor, this cofactor also has an influence on the three-dimensional structure and stability of the molecule. This refers, in particular, to the prosthetic heme group in myoglobin and other proteins. We cannot yet estimate theoretically the stability of the globule towards denaturation.

As has already been said (page 119), the problem of the correspondence between the primary and the three-dimensional structure is directly tied up with the foundations of theoretical biology. As a matter of fact, two problems are involved here: the correspondence between the primary and the three-dimensional structure of the protein and the relationship between the three-dimensional structure of the protein and its biological function.

The correspondence between the primary and the three-dimensional structure is not unambiguous—to one and the same three-dimensional structure of the protein there may correspond different amino-acid sequences. This has been demonstrated, for example, for globins—proteins that contain heme groups, which store and transport oxygen. The primary structure of globins is strongly varied. But globins have similar three-dimensional structures—they contain practically the same α-helical portions and "pockets" that accommodate heme groups.

It may be thought that the biological function of a protein is determined by its three-dimensional structure. The extent of unambiguity of this correspondence is not clear as yet; presumably, it is different for different proteins. Anyway, there is every ground for believing that changes in the primary structure of a number of proteins provoked by mutations may not affect the biological functions of these proteins. In other words, mutations at the molecular level may prove to be neutral or nearly neutral for natural selection. This forms the basis for the "*neutralistic*" *theory of evolution* advanced by Kimura in 1968. By analysing the evolution of homologous proteins, Kimura came to the conclusion that first, for a particular protein the rate of evolution, expressed by the number of amino-acid replacements per geological unit of time, is nearly constant and is the same for different evolutionary sequences. Second, the molecules of proteins or their constituent parts, being subject to relatively small functional constraints, undergo evolution faster than molecules characterized by strong constraints. For example, the numbers of amino-acid replacements per amino-acid residue for 10^9 years are as follows: 9.0 for fibrinopeptides; 1.4 for the chains of hemoglobin; 0.3 for cytochrome *c*; 0.006 for histone H4.

Thus, molecular evolution predominantly involves neutral mutations, the random drift of genes. Natural selection operates at the phenotype level rather than at the level of biopolymer molecules. The "neutralistic" theory met with sharp objections as contradicting

the modern Darwinian theory of evolution. As a matter of fact, no contradiction exists between "neutralism" and "selectionism". The physical essence of the "neutralistic" theory consists in the ambiguity, degeneracy of the correspondence between the primary and the biologically functional (three-dimensional) structure of the protein.

4.7. The Structure and Stability of the Globule

The following sources of information on the structure of the protein globule are available at present:

1. The X-ray diffraction analysis, which in an ideal case enables one to locate each atom in the protein molecule (see Sec. 5.1).
2. The study of the scattering of light and X-rays (Sec. 5.2), which furnishes data on the shape of the globule as a whole.
3. Optical methods, optical spectroscopy and NMR and EPR spectroscopy (see Secs. 5.4 through 5.10), which give information on the fraction of the residues present in the α-helical, *cross* β- and disordered forms, their interactions and mobility, etc.
4. The study of *deuterium exchange* (the exchange of the hydrogen atoms of a protein in heavy water for deuterium atoms), which allows one to estimate the fraction of residues linked by hydrogen bonds and to identify the internal and external atomic groups in the protein.
5. The study of electrophoresis, which gives information on charges on the globule.
6. The study of denaturation and proteolysis which yields data on the stability of the globule. Here the information obtained by means of *microcalorimetry* is especially important.
7. The results of theoretical calculations and estimates.

As examples of globular proteins we shall consider myoglobin (Mb) and hemoglobin (Hb). The structure of these proteins has been studied in detail by X-ray diffraction methods (Kendrew and Perutz, respectively). Figure 4.18 shows the structure of myoglobin. The myoglobin molecule, which contains the prosthetic heme group (marked by an arrow; compare with pages 53 and 122), is built up of 8 alpha-helical portions intersected by disordered portions. Approximately 75 percent of the total number (153) of residues belong to these portions, which are marked by the letters A to H (counting from the N-terminal end of the chain); the residues within each α-helical portion are numbered from 1 to *n*. The interhelical, disordered portions are designated AB, CD, etc., and the terminal portions are denoted by NA and HC. In contrast to myoglobin, the hemoglobin molecule has a quaternary structure: it is composed of four chains coiled into globules—two chains designated as α and two chains as β. The structure of each of the α- and β-globules is similar to the structure of myoglobin.

A detailed analysis of the structures of myoglobin and hemoglobin shows that the cores of their globules are really filled with tightly packed nonpolar side groups R of the amino-acid residues. The number of inner residues in each of the four subunits of hemoglobin is 36. Many of the Gly and Ala residues, while being slightly hydrophobic, are arranged on the surface of the molecule. Some bulky nonpolar side chains, which are outside the globule, are hidden in the cavity

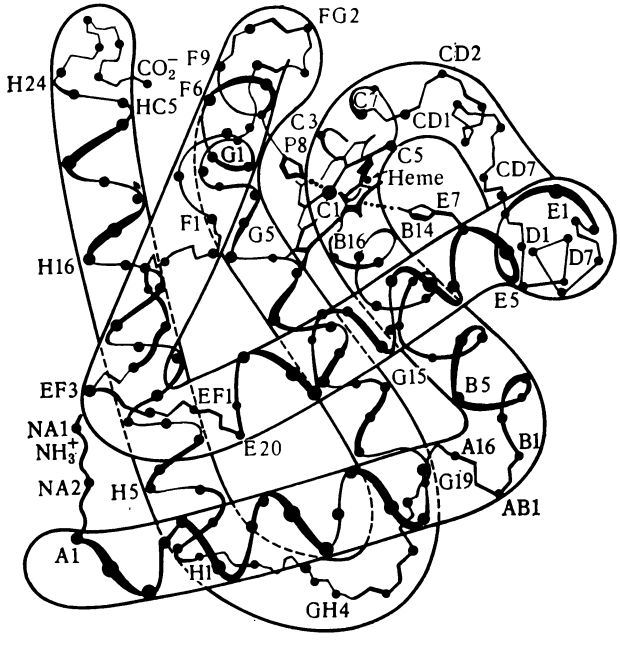


Fig. 4.18. The structure of myoglobin.

near the surface, which reduces to a minimum their contacts with water. All the side chains that are ionizable at neutral pH are present on the surface of the globule. The same is valid for other polar side chains, with the exception of heme-bound His and Thr C4 which are linked by a hydrogen bond. On the whole, of the 77 polar groups in myoglobin only 5 or 6 groups are inside the globule, the remaining ones residing on its surface. The study of the hemoglobins of various species of vertebrates and of sperm whale and human myoglobins has demonstrated that upon replacement of 33 inner residues their

nonpolar nature is retained. There are only 10 invariantly nonpolar residues on the surface of the globule. The analysis of the α-helical portions A, B, E, G, and H points to the periodic arrangement of the nonpolar residues in them. One side of the α-helix bears hydrophobic residues and the other predominantly hydrophilic ones. The hydrophobic side is oriented towards the core of the globule.

All these facts confirm what has been said earlier (Secs. 4.5 and 4.6). Hydrophobic interactions play a most important role in the self-assembly of the globule. The secondary structure—the α-helical portions (and also those portions that are in the β-form)—are also determined by hydrophobic interactions.

The electrostatic salt bridges in proteins are also stabilized by a water environment, since their formation leads to the release of oriented water molecules that surround the charged groups. Thereby, the formation of a salt bridge is accompanied by an increase in the entropy of water. This gain in free energy is more considerable than that provided by the electrostatic attraction of charges. The effect of water on the salt is, however, different from the hydrophobic effect—the salt bridges become stronger, whereas the hydrophobic interactions are weakened upon addition of nonaqueous solvents.

The apparent density of proteins in water is higher than their dry density in organic solvents. This increase in density is caused by the *electrostriction* of the bound water. The water molecules are bound on the globule surface and also in the interior of the globule—between the domains of chymotrypsin, for example, or between the subunits of a protein having a quaternary structure. The amount of bound water, into which no electrolytes can gain entrance, is about 0.3 g per gram of protein, i.e., approximately 100 water molecules per protein with a molecular weight of 6000. The fact that electrolytes cannot get into bound water is accounted for by electrostatic effects. Let us consider a charge e immersed in a solvent of high dielectric constant ε_0, which covers a layer of a substance of low dielectric constant ε_s. In this layer there arises an induced charge equal to

$$e' = e\frac{\varepsilon_0 - \varepsilon_s}{\varepsilon_0 + \varepsilon_s} \tag{4.42}$$

Accordingly, the energy of the charge e at a distance r from the nonpolar layer is equal to

$$E = \frac{ee'}{2\varepsilon_0 r} = \frac{\varepsilon_0 - \varepsilon_s}{\varepsilon_0 + \varepsilon_s}\,\frac{e^2}{2\varepsilon_0 r} \tag{4.43}$$

If $\varepsilon_0 = 81$, $\varepsilon_s \approx 5$, then $E = kT$ at $r = 0.3$ nm, i.e., ions cannot penetrate into the aqueous layer of about monomolecular thickness on the protein.

One can naturally ask: How is the theory of the polymeric globule (see Sec. 3.5) related to the structure and properties of the real

protein globule? The Lifshitz theory is the theory of homopolymers. It shows that there are possible phase transitions from a random coil to a globule and from a globule to a random coil and that the globule is a condensed crystalline system. Both these propositions correspond to the properties of proteins. Of importance for the three-dimensional structure of the protein is, however, its heteropolymeric structure, the variety of residues and their interactions. Protein chains are short as compared with those dealt

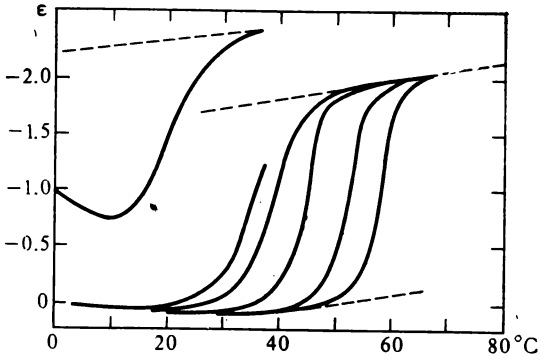


Fig. 4.19. The dependence of the extinction coefficient at 293 nm of a solution of chymotrypsinogen on temperature. The curves were recorded at different pH values. The lower dashed curve corresponds to the native form and the upper line to the denatured form.

with in the Lifshitz theory. Therefore, the protein globule has no fluctuating fringe and the phase transition from a protein globule to a coil is a first-order transition.

This transition occurs during the denaturation of the protein. Denaturation, which is the destruction of higher structural levels with retention of the primary structure of the chain, is caused by the action of acids, alkalis, and substances such as urea, by a mechanical effect (*ultrasound*), by heating, etc.

Figure 4.19 shows the results of a study of the thermal denaturation of chymotrypsinogen. The transition manifests itself in a change in the extinction coefficient at 293 nm (see page 153). The process is reversible and the curves are at equilibrium. Such a transition is typical for globular proteins. The temperature range of the transition is of the order of 10°C. The transition occurs in a single stage, like a first-order phase transition.

Calorimetric investigations furnish information on the thermodynamics of protein denaturation. Privalov has shown that the absorp-

tion of heat when certain globular proteins (albumins, myoglobin, chymotrypsinogen, ribonuclease) are heated occurs in two steps. The first step is predenaturation characterized by a certain increase in heat capacity without a jumpwise increase in enthalpy, and the second step is denaturation proper involving a phase transition. The enthalpy and temperature of denaturation depend strongly on the pH of the medium, which points to a considerable role of electrostatic effects. For example, myoglobin at pH = 12.2 has a melting point of 50°C and a value of ΔH equal to 300 kJ/mole of protein; at pH = 10.7 the melting point is 78°C and the value of ΔH is 710 kJ/mole of protein. The order of the free-energy change ΔG upon denaturation of a globular protein is 4.2-6.3 kJ/mole, that of ΔH is tens and hundreds of kJ/mole, and of ΔS is 700-800 J/(mole·K).

Let us now describe the method of determination of ΔG from denaturation in an urea solution (Tanford). For this process we have

$$-\Delta G_u = RT \ln K_u = RT \ln \frac{[D]}{[N]} \tag{4.44}$$

where [D] and [N] are the concentrations of denatured and native proteins. We are interested in the values of ΔG_{H_2O} for an aqueous solution. We make use of the cycle

$$\begin{array}{ccc} \text{Native protein in water} & \xrightarrow{\Delta G_{H_2O}} & \text{Denatured protein in water} \\ \Big\downarrow \Delta G_N & & \Big\downarrow \Delta G_D \\ \text{Native protein in urea solution} & \xrightarrow{\Delta G_u} & \text{Denatured protein in urea solution} \end{array}$$

Then

$$\delta\Delta G = \Delta G_u - \Delta G_{H_2O} = \Delta G_D - \Delta G_N \tag{4.45}$$

The quantity $\delta\Delta G$ can be represented in the following form:

$$\delta\Delta G = \sum_i y_i n_i \Delta G_i \tag{4.46}$$

where n_i is the number of groups of type i in the protein; ΔG_i is their contribution to the free energy of the transition; y_i is the numerical parameter which depends on the extent of accessibility to the solvent of groups of a given type in the native conformation. The agreement with experiment is obtained if the values of y_i for polar and hydrophobic groups and peptide bonds are assumed to be equal to 0.25, 0.75, and 0.50, respectively. The values of ΔG_i are found from data on the solubility of amino acids in water and in a urea solution of a given concentration.

Let c_u^* be the concentration of urea at which [D] = [N], i.e., the point of half-denaturation. Then, according to Eq. (4.44), $\Delta G_{u_i}^* = 0$

and

$$\Delta G_{H_2O} = -(\delta \Delta G)^*_{c_u} = -\sum_i y_i n_i (\Delta G_i)^*_{c_u} \tag{4.47}$$

This is a convenient method of determining the conformational stability of the protein.

The free energy of denaturation must be the sum of many contributions. We can write

$$\Delta G = \Delta G_\alpha + \Delta G_H + \Delta G_{np} + \Delta G_e + \Delta G_s + \Delta G_b + \ldots \tag{4.48}$$

where ΔG_α is determined by α-helix-coil transitions, ΔG_H by the breaking of hydrogen bonds between the protein chains of neighbouring macromolecules, ΔG_{np} (the subscript np stands for nonpolar) by a change in hydrophobic interactions, ΔG_e by a change in electrostatic interactions, ΔG_s by the swelling of the coil formed, and ΔG_b by the breaking of weak bonds between portions of the secondary structure. We have

$$\Delta G_\alpha = (N - 4)\, \Delta H_\alpha - (N - 1)\, T \Delta S_\alpha \tag{4.49}$$

where N is the number of units in the α-helix; ΔH_α and ΔS_α are the enthalpy and entropy changes per chain link with $N \gg 1$ during a helix-coil transition. A chain consisting of N links has, in an α-helical form, $N - 4$ hydrogen bonds, which restrict the mobility of $N - 1$ links. According to the estimates available in the literature, $\Delta H_\alpha = 6.3$ kJ/mole and $\Delta S_\alpha = 17.6$ J/(mole·K). If only ΔG_α and ΔG_s are taken into account, the transition temperature can be found from the condition $\Delta G = \Delta G_\alpha + \Delta G_s = 0$, i.e.,

$$T_m = \frac{(N-4)\,\Delta H_\alpha + \Delta H_s}{(N-1)\,\Delta S_\alpha + \Delta S_s} \tag{4.50}$$

Here the value of ΔH_s may be neglected. The quantity ΔS_s is proportional to N—the longer the chain, the larger is the number of cross-links. The estimation of ΔG_{np} gives 2.1 kJ/mole and the values of ΔG_H and ΔG_e are of the same order. A rigorous theoretical evaluation of all the contributions to ΔG is difficult.

The fraction of denatured material can be given by the following formula:

$$x = \frac{\exp(-\Delta G/RT)}{1 + \exp(-\Delta G/RT)} \tag{4.51}$$

At the transition point $\Delta G = 0$ and $x = 0.5$. The sharpness of the transition is characterized by the derivative

$$\left(\frac{dx}{dT}\right)_{Tm} = \frac{\Delta H}{4RT_m^2} \tag{4.52}$$

The experimental and theoretical study of the denaturation of proteins provides information, on the one hand, about the degree

of stability of native protein (the free-energy change) and, on the other, about the cooperativity of interactions which determines the sharpness of the transition. Of course, the investigation of denaturation does not elucidate the organization of the protein globule. Denaturation means the destruction of this aperiodic crystal. Nonetheless, as a result of such destruction, we obtain averaged information on the stability of the internal bonds and on the *electronic-conformational interactions* (see Sec. 6.5).

Nasonov and Aleksandrov (1940) presumed that the functioning of proteins is associated with their "partial denaturation". Later, Aleksandrov established correlations between the temperature of the body and the temperature of the denaturation of its proteins. This is illustrated, for example, by Table 4.6, which compares the respective characteristics of two related species of frogs—the more Northern brown frog (*Rana temporaria L.*) and the more Southern lake frog (*Rana ridibunda Pall.*).

Table 4.6. The Thermophilicity of Frogs (according to Aleksandrov)

Characteristic	Common frog	Lake frog
The boundary of the area of distribution:		
Northern	70° North	60° North
Southern	43° North	40° North
Upper	3048 m	2438 m
The body temperature in nature (°C)	6.0-26	11.0-29.5
The temperature chosen (°C)	13-26	18-28
Loss of reflex excitability of the organism at temperature, °C	30-32	35-36
Loss of the mobility of spermatozoa	39.2	41.4
Loss of the ability to develop ovum	36.6	38.5
Complete inactivation of ATPase activity of muscle homogenates	40-42	46-50
Hemoglobin, 50-percent denaturation	60	64
Tendon collagen, 50-percent contraction	51.5	57.6
Skin tropocollagen, 50-percent denaturation	25.5	32.0
Serum albumin, start of coagulation	68.0	75.0

The temperature of protein denaturation by far exceeds the body temperature. Aleksandrov explained this correlation by the fact that the functioning of the protein requires a certain level of mobility, internal flexibility of the globular or fibrous protein molecule (for the importance of this mobility to enzymic catalysis, see Chapter 6). The lower the body temperature, the weaker must be the intramolecular interactions responsible for the rigidity of the molecule and the lower must be the temperature of denaturation. It is possible that here an especially important role is played by interactions taking place in the globule. Probably, electrostatic interactions (salt bridges)

are also of importance. A remarkable example is provided by the properties of a small protein, *ferredoxin*, which contains about 60 amino-acid residues and two Fe_4S_4 clusters. Ferredoxin from the bacterium *Clostridium pasteurianum* is completely inactivated when heated up to 70°C for two hours; ferredoxin from *C. acidiurici* retains upon such heating 20 percent of activity, and those from *C. tartarivorum* and *C. thermosaccharolyticum* retain 50 and 90 percent of activity, respectively. It has been found that the increased heat stability of a protein from thermophilic organisms is due to the extra salt bridges. One such bridge can raise the free energy of stabilization up to 12-16 kJ/mole.

Well correlated with denaturation is proteolysis, i.e., the hydrolytic scission of a protein with the aid of proteolytic enzymes (Linderstrom-Lang). The higher the thermal stability of the protein, the more difficultly it breaks down. This is important, in particular, for frogs—proteolysis of collagen with collagenase occurs when the tadpole loses its tail.

The denaturation of a protein by an acid or a base is directly associated with electrostatic interactions. The free energy of a polyelectrolyte coiled into a globule is proportional to the square of its surface charge, since the stability of the protein decreases on both sides of the isoelectric point (page 32). Besides, the titration with an acid or a base can ionize the internal groups, which enhances the attraction of the hydrate shell and shifts the equilibrium to the side of a coil.

4.8. Antibodies and Antigens

The protection of an organism from foreign biopolymers and, thereby, from infectious microorganisms, is accomplished through *cellular* and *humoral immunity* (see Sec. 17.9). In the second case, immunity is determined by the interactions of *antibodies*—specific proteins produced by lymph cells—with foreign biopolymers, which are called *antigens* in this case. The immune response, i.e., the appearance of antigens in the organism, is the result of recognition of antigens by certain populations of lymphocytes. The process occurs at the level of the organism and involves various cellular recognizing systems, which are in the process of "learning" since they become capable of remembering the antigen introduced for the first time and can respond to its repeated injection by an enhanced production of antibodies.

Antibodies belong to a class of serum proteins known as immunoglobulins. A human being has five principal classes of immunoglobulins denoted as IgG, IgM, IgA, IgD, and IgE. The immunoglobulins IgG function as an antibody. Their molecular weights are about 150,000 and the sedimentation coefficient is 7 S. The immunoglobu-

lins IgM are the pentamers of IgG. The structure of many IgG has been studied. Figure 4.20 shows schematically the structure of rabbit IgG antibodies. The molecule consists of two heavy chains with a molecular weight of 53,500 ±•4000 (about 450 residues) designated as γ and two light chains with a molecular weight of 23,800 ± 1000 (about 220 residues) designated as κ or λ. The chains are joined by disulphide bridges.

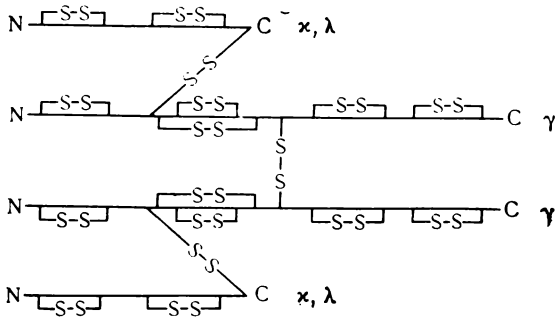


Fig. 4.20. Schematic representation of the structure of the antibody IgG.

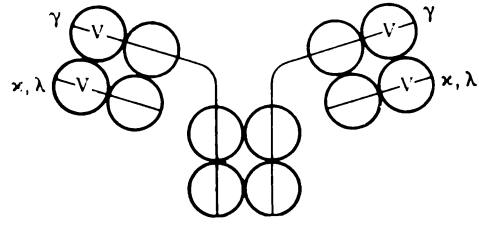
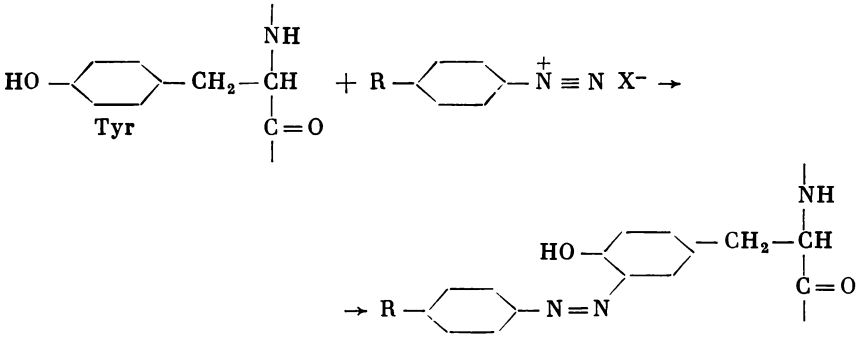


Fig. 4.21. Schematic of the domain structure of IgG according to Edelman.

Antibodies to the various antigens have similar structures, despite the differences in the primary structure localized in the variable region (V-region) of the molecule which lies in the vicinity of the N-terminal ends of the light and heavy chains. Presumably, antibodies have a domain structure consisting of several interlinked globules (Fig. 4.21). The variable regions are responsible for antibody-antigen interactions, since the antibody specificity is controlled by the specific features of the primary structure. The active centre of antibodies contains a variable region. There are two such centres in each antibody molecule.

Antigens are primarily proteins and polysaccharides. Landsteiner (1919) worked out a method for synthesis of artificial antigens which

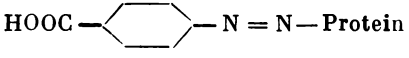
is based on the combination of diazotized aromatic compounds with the Tyr of the protein:



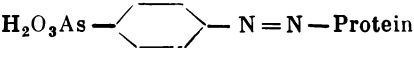
Thus, any radical R can be introduced into the protein. The R-phenylazo protein can be used as an antigen for stimulating the generation of respective antibodies.

It has been established that the factor that determines antigenic specificity is just the radical R, and the protein plays a secondary role. The antibody produced in response to R-P (P stands for a protein) reacts with R-P′ (P′ stands for another protein) but not with R′-P. The antigen R-P produces with a respective antibody a difficultly soluble precipitate. If we add to the R-antigen-R-antibody system small molecules containing the same group R, the antigen-antibody reaction is retarded, and when the concentration is further increased, the reaction ceases to proceed. Small molecules do not generate antibodies in the organism and are not antigens. They interact, however, with the earlier generated antibodies, forming soluble compounds. This is the so-called hapten action, and the small molecules mentioned are known as *haptens* or *determinant* groups. Haptens compete with antigens, interacting with the same active regions of antibodies. The reaction between antibody and hapten or antigen takes place through weak interactions but it obeys the mass action law. Natural antigens are multivalent, i.e., they contain several determinant groups (haptens). The three-dimensional structure of the determinant group is essential for immunological processes.

The presence of reactive groups (active centres or sites) in antibodies is proved by the elegant experiments carried out by Pressman and Sternberg (1951). A rabbit was immunized by artificial synthetic antigens that contained the *p*-azobenzoic acid residue



or the *p*-azophenylarsenic acid residue



as determinant groups. The resulting antigens were subjected to iodination, as a result of which their immunological activity was completely suppressed. However, when iodination was carried out in the presence of haptens, the activity was retained. Evidently, the introduction of iodine into the active region of antigen abolishes its activity, and the hapten protects this region against the iodine.

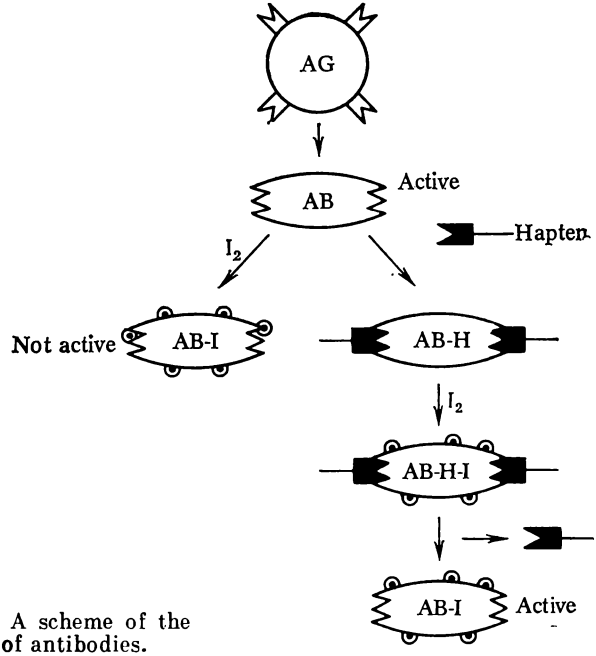


Fig. 4.22. A scheme of the iodination of antibodies.

What has been said is depicted in Fig. 4.22. There is a structural correspondence between the active region of antibody, on the one hand, and the hapten or determinant group of antigen, on the other.

The structure of a complex of antigen with the corresponding bifunctional dinitrophenyl hapten (DNP) has been studied by means of electron microscopy.

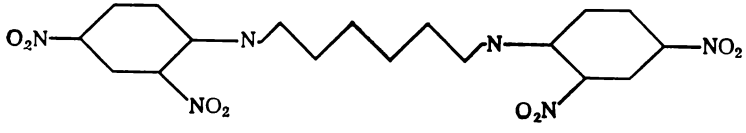


Figure 4.23 shows an electron microphotograph of the antibody-DNP complex and Fig. 4.24 is its schematic interpretation. Divalent

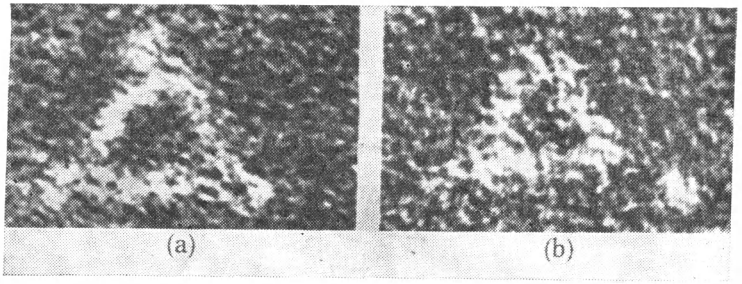


Fig. 4.23. Electron micrographs of an antigen-antibody complex (*a*) and of the same complex treated with pepsin (*b*).

antibodies, interacting with divalent DNP molecules, form ternary cycles. The projections at the corners of the triangles are portions of antibodies that do not contain variable regions. When antibody is preliminarily treated with pepsin, these portions are split off (Fig. 4.23*b*).

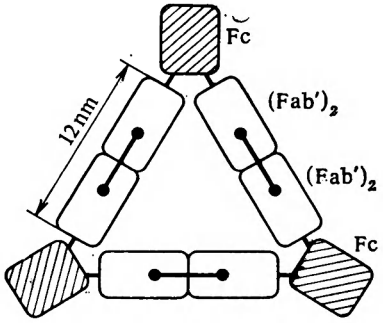


Fig. 4.24. Schematic representation of an antigen-antibody complex. The "dumbbell" is the dinitrophenyl molecule.

The equilibrium constants for the antibody-antigen reaction or antibody-hapten reaction are determined by carrying out experiments *in vitro*:

$$K = \frac{[\mathrm{AB}]-[\mathrm{H}]}{[\mathrm{AB}][\mathrm{H}]} = \exp\left(-\frac{\Delta H - T\Delta S}{RT}\right) \tag{4.53}$$

Here AB stands for antibody and H for hapten. The dependence of K on $1/T$ gives the values of ΔH and ΔS. The enthalpy changes are of the order of 10 kJ/mole (weak interactions); the changes ΔS are, as a rule, positive, which may be accounted for by the liberation of hydrating water molecules during the formation of complexes.

The investigations of the dependence of the equilibrium constants

on the structure of the hapten or of the corresponding determinant group of antigen have been carried out by Pauling and Pressman (1944-1957). In particular, there were obtained antibodies to the ions of *ortho*-, *meta*-, and *para*-azobenzene arsonate. Antibodies to each of these groups function as if they were structurally complementary with the van der Waals surface of the determinant group. The dependence of the interaction on the structure is strongly pronounced. For example, the antibody to *ortho*-azobenzene arsonate binds well other *ortho*-substituents (with the radicals OH, NO_2, Cl, CH_3, NH_2 instead of AsO_3H^-), less strongly *meta*-substituents and most weakly *para*-substituents. Very important is the relative arrangement of parts of the hapten. For antibodies against *para*- $-N=N-C_6H_4-AsO_3H^-$, the relative constants for binding the haptens $OAsO_3H^-$, $H_3CAsO_3H^-$, $H_5C_6AsO_3H^-$, and $H_5C_6CH_2AsO_3H^-$ are, respectively, 0, 0, 1, 0. The removal of the benzene ring (which is needed for binding) from the arsonate by one CH_2 group destroys the binding. The binding depends strongly on the nature of the charged group. For example, the same antibodies strongly bind $C_6H_5AsO_3H^-$ and $C_6H_5PO_3H^-$, but they do not bind $C_6H_5SO_3^-$, $C_8H_5CO_2^-$ and $C_6H_5SbO_5H_4^-$. The ionic size is also important.

These and other data point to complementarity, the structural fit between the active site of antibody and the determinant group of antigen or hapten. The study of the interaction of antibody to poly-L-alanine with alanine oligopeptides has allowed the sizes of the active site of the antibody to be determined: $2.5 \times 1.1 \times 0.2$ nm^3. Antibodies are globular proteins. Upon denaturation of antibodies their ability to bind specifically antigens or haptens disappears. In some cases it is possible to accomplish the renaturation of antibodies with their activity being recovered. The presence of hapten favours renaturation.

Molecules of antibodies exhibit some flexibility, i.e., the ability to undergo conformational changes. The polarized luminescence of complexes of IgG with luminescent dyes has enabled the times of rotational relaxation τ to be determined; they proved to be of the order of 50 ns (see Sec. 5.5). These values correspond to the Brownian rotational motion not of the entire protein molecule but of its small portions, i.e., are indicative of the flexibility of the protein molecule. Presumably, the domains exhibit mobility. The interaction of hapten with antibody leads to a noticeable increase in τ, which is an indication of the change of the antibody conformation. It has been found that the conformation of antigens is also changed when an antibody-antigen complex is formed. The data of optical measurements are confirmed by investigations of EPR spectra of antibodies that contain paramagnetic labels.

4.9. Fibrous Proteins

In concluding this chapter we shall dwell on the structure and properties of fibrous proteins—structural and contractile proteins. The former play the role of supporting and protective components and are found in tendons, cartilage, bone, ligament, etc. (*collagens*), and also in skin, hair, wool, horn, etc. (*keratins*). The contractile proteins are the working substances of contraction systems, in particular, of muscles (*myosin*).

In contrast to the majority of globular proteins, fibrous (or fibrillar) proteins function not in solution but form supermolecular tissue systems.

If a connective tissue is extracted by cold solutions of neutral salts at the physiological value of ionic strength, then a part of collagen will be dissolved. The other part, called *procollagen*, dissolves in citric or acetic acid at pH = 3.8.

Collagen fibres (fibrils) are produced upon aggregation of procollagen chains. These fibrils play the role of centres of crystallization during osteogenesis (the growth of bones).

Collagenic fibres are insoluble in water, but on prolonged heating with water collagen is converted into soluble gelatin as a result of the degradation of the long protein chains.

The amino-acid composition of collagen is specific. The 33 percent of all the amino acid residues are Gly, 12 percent are Pro, and 10 percent are the noncommon residue *hydroxyproline* (Hyp). This residue and also *hydroxylysine* (Hyl), whose content ranges from 0.3 to 1.2 percent, is found mainly in collagens. Collagen contains also up to 10 percent of Ala and considerably lower amounts of other amino acids, the contents of Tyr, His, Cys, Met, Val, and Phe being especially low, less than 1 percent each. Thus, up to two-thirds of the residues are Gly, Pro, Hyp, and Ala. The amino-acid sequence in the collagen chain may be depicted as follows: $(\text{Gly-X-X})_n$, where X is any residue. The following sequences are most frequently encountered: Gly-Pro-Hyp, Gly-Pro-Ala, and Gly-Ala-Hyp.

The structure of collagen has been established in the X-ray diffraction studies carried out by Ramachandran, Rich, and Crick. The collagen molecule in solution, called *tropocollagen*, is built up in the form of a triple helix with a molecular weight of about 360,000, each of the three chains containing about 1000 amino-acid residues. The length of the molecule is 290 nm and its diameter 1.5 nm.

A theoretical calculation of the structure of tropocollagen is based on the minimization of conformational energy and takes account of the stabilizing role of water (Tumanian). This calculation leads to results which are in agreement with experiment. Tropocollagen experiences a thermal transition—denaturation—in solution. Depending on the species of vertebrates, the temperature of denaturation, T_d, varies from 20 to 40°C; the enthalpy of denatura-

tion ranges from 3150 to 6300 J/mole and the entropy from 10.9 to 21 J/(mole·K) (calorimetric data). The values of T_d, ΔH, and ΔS increase with increasing content of amino-acid residues (Pro, Hyp), which do not form intramolecular hydrogen bonds C=O ···· ···· H—N. Presumably, tropocollagen is really stabilized by neighbouring water molecules, which form hydrogen bonds with the imine nitrogen atoms: N ···· H—O. The removal of water leads to the destruction of the structure of collagen, which is regained up on further addition of water.

The problem of the relationship between the properties of collagen and its structure cannot yet be considered to have been solved completely. Work in this field is necessary—collagen is one of the most important proteins and, having a simple composition and structure, it serves as a valuable model for the study of all proteins.

Let us turn to *keratin*. The macroscopic properties of hair, wool, etc., point to its high stability and insolubility. These properties are determined mainly by a large number of disulphide cross-links between polypeptide chains. The keratin of human hair and wool keratin contain 11-12 percent of cystine residues, i.e., 3 percent of sulphur.

The length of keratin fibres is essentially dependent on the content of water in them (the hair hygrometer is based on this fact). These fibres are elastic and extensible. The pioneering works devoted to the X-ray diffraction studies of proteins were those of Astbury, who studied the structure of wool keratin (1943). Keratin in its natural unstretched state is characterized by a periodicity or repeat unit of 0.51 nm (α-keratin); on stretching, it is converted into what is known as β-keratin with a repeating spacing along the fibre axis of 0.33 nm and with 0.47 and 0.97 nm spacings at right angles to this axis. Astbury maintained that α-keratin exhibits regular folds of chains, which straighten out on stretching. Later, it was shown that α-keratin is composed, to a significant extent, of α-helices, and β-keratin has a canonical β-form.

Keratin is a complex protein. When the disulphide linkages are ruptured, which occurs on oxidation or reduction, there is produced a soluble substance, from which two fractions can be isolated, one rich and the other impoverished in sulphur. The former consists of fibrous and the latter of globular molecules. Probably, globular molecules serve as cross-links in keratin fibres. The structural unit of the fibre is a cylindrical microfibril having a diameter of about 7.5 nm and made up of proteins with a low content of sulphur. Figure 4.25 shows a hypothetical model of the fibre of α-keratin. The regularly arranged portions are α-helices which are twisted in pairs. The protein is heterogeneous and is composed of two main components in the ratio of 2 : 1. The individual molecules may be arranged either sequentially (Fig. 4.25*b* and *c*) or in parallel to

each other (Fig. 4.25*d*). Experiment gives large repeating spacings along the fibre axis equal to 20 nm, which is in harmony with the models shown in Fig. 4.25*c* and *d* but not with those of Fig. 4.25*b*.

The microfibril as a whole is made up of 11 such protofibrils—double (and, maybe, also triple) helices; two of these occupy the

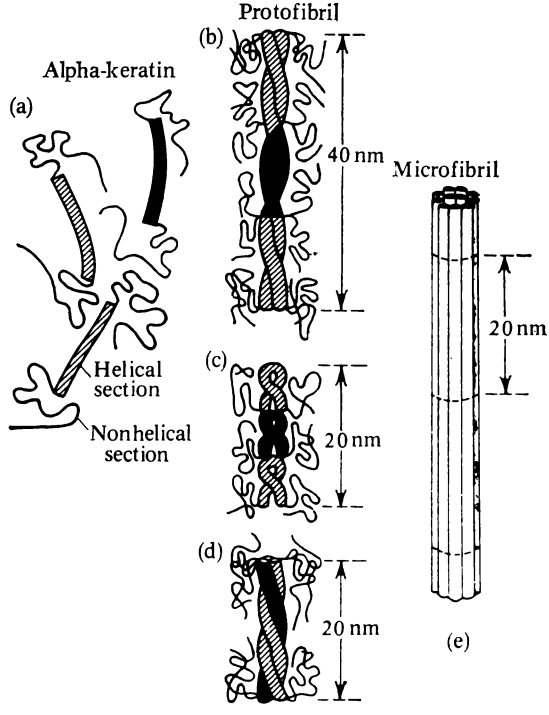


Fig. 4.25. A hypothetic schematic representation of the molecular organization of α-keratin:

(a) The individual molecules of the two main components (black and hatched; the latter are proteins with a low sulphur content); (b), (c), and (d)—protofibrils; (e)—microfibril.

position in the centre of the microfibril and nine at the periphery (Fig. 4.25*e*). The 9 : 2 ratio is characteristic of a number of biological fibrillar structures (see Sec. 12.6).

The microfibril has a homogeneous structure. The lengths of helical portions that form protofibrils are nearly equal for different species, but the composition of nonhelical portions varies significantly. The differences in the keratins of the vertebrates are determined by the way in which the microfibrils are packed and by the com-

position of the sulphur-rich globular component. β-Keratin has been studied to a much lesser extent.

Mention should be made here of another fibrous protein—*silk fibroin.* It consists predominantly of Gly (more than 40 percent) and also Ala, Ser, and Tyr. The amino acids cystine and methonine are absent, the remaining amino acids being present in small amounts. The X-ray diffraction patterns of silk fibroin are similar to those of β-keratin, the main conformation being the β-form. The silk of *Bombyx mori* is made up of the polypeptide $(Gly\text{-}Ala)_n$, and the silk of *Lyssax* is composed of polyalanine.

Contractile proteins will be discussed in Chapter 12.

Chapter 5 Methods of Investigation of the Structure of Biopolymers

5.1. X-Ray Diffraction Analysis

Before we proceed further with proteins we shall consider the methods of investigation of the structure of biopolymers.

Natural sciences make use of two principal methods for the study of the structure of atoms and molecules. These methods are chemistry and optics in the wide sense of the word, i.e., the study of the interaction of a substance with light over the entire range of wavelengths—from X-rays to radio waves. Chemistry discloses the primary structure of biopolymers, the structure of the functional centres of protein globules, etc. Chemistry as such cannot, however, establish the spatial three-dimensional structure of proteins or nucleic acids.

X-ray diffraction gives direct information on the arrangement of atoms in molecules and crystals. X-rays, i.e., electromagnetic radiation with a wavelength of the order of 0.1 nm, are scattered by the electron shells of atoms. The interference of waves scattered by a substance produces a diffraction pattern. If X-rays are scattered by a crystal, the diffraction may be regarded as the reflection of X-rays from the planes of the crystal lattice (Fig. 5.1). The diffraction of X-rays is observed if the scattered waves are in phase, i.e., the difference in distance travelled by rays is equal to an integer n of wavelengths. If the distance between the reflecting crystalline planes is d (the interplanar spacing), the condition for diffraction (reflection) is given by the Bragg equation

$$n\lambda = 2d \sin \theta \qquad (5.1)$$

where λ is the wavelength and θ is the angle between the direction of the incident light ray and the crystalline plane. The derivation of the equation is clear from Fig. 5.1.

Diffraction is determined by the fact that the interatomic distance d in the lattice (0.1-0.4 nm) is of the same order of magnitude as the wavelength (use is often made of the K_α radiation of Cu with $\lambda = 0.154$ nm). The main idea of X-ray diffraction analysis consists in determining the spacing d on the basis of a diffraction pattern obtained for rays of known wavelength. Detailed information on the

structure of a substance can be obtained only through the investigation of crystals, in which the atoms are arranged in a regular periodic manner. Here, of course, the elementary formula (5.1) is insufficient. The crystal lattice is characterized not by one but by three

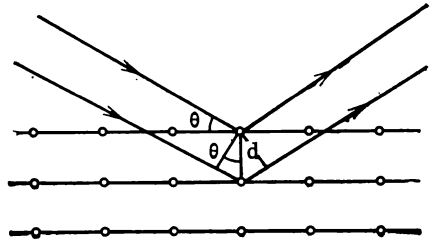


Fig. 5.1. X-ray diffraction.

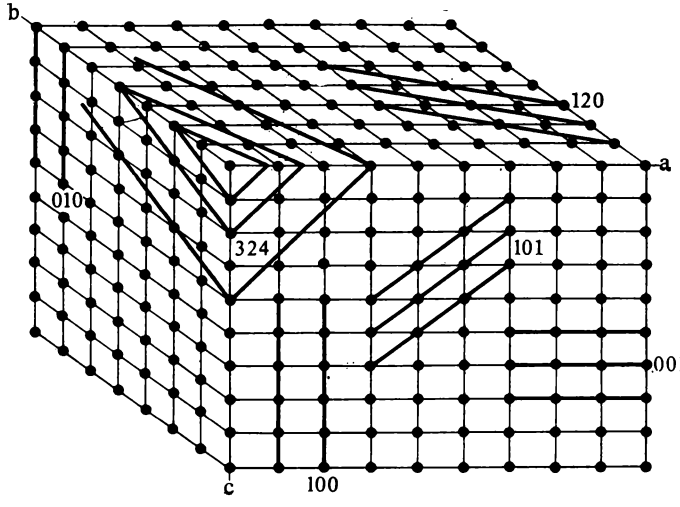


Fig. 5.2. Sections of the reflecting planes in the orthogonal crystal lattice.

periods and in a crystal there is a set of crystal planes, from which X-rays are reflected. These planes can be characterized by discrete integers—*Miller indices*; the interplanar spacings d are different for different indices. Figure 5.2 shows the intersections of different reflecting planes in the orthogonal crystal lattice. Accordingly, on an X-ray diffraction pattern (a *Laue pattern*) produced by means of monochromatic radiation there is observed a series of diffraction maxima. Their number is especially large for crystals of proteins, which is accounted for by relatively large periods of the crystal

lattice. Figure 5.3 shows the X-ray diffraction pattern of sperm whale myoglobin. An X-ray diffraction pattern is usually recorded on photographic film, but ionization methods of recording are much more accurate and sensitive.

Diffraction maxima are of various intensities. Analysis of diffraction patterns enables one to derive the electron-density distribution in a crystal. The electron density is expressed through so-called structure amplitudes, which depend on the number of electrons in the atom and on the direction of scattering. Structure amplitudes are complex quantities characterized by moduli and phases. Without knowing the phases it is impossible to establish the structure of the material. The method of determination of phases worked out by Perutz for proteins consists in that to molecules that make up a crystal there are added heavy atoms, say mercury atoms. A heavy atom, which has many electrons and which strongly scatters X-rays, brings about appreciable changes in the intensity of diffraction spots. The phase can be determined from the difference in amplitudes in the presence and in the absence of a heavy atom. The use of protein derivatives containing several heavy atoms allows one to solve the phase problem unambiguously. In the heavy-atom method it is necessary that the structure of the protein crystal remain intact upon introduction of a heavy atom. The Perutz method is the *method of isomorphous replacements*—the mercury derivatives of a protein are crystallized isomorphously with the unsubstituted protein.

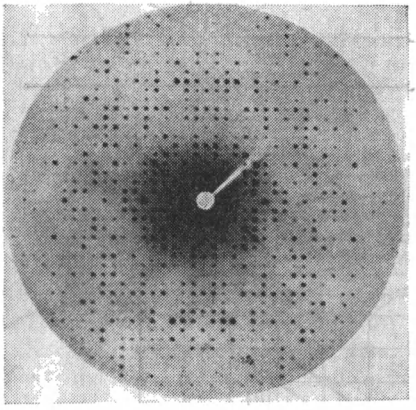


Fig. 5.3. The X-ray pattern of the sperm whale myoglobin.

The electron-density distribution $\rho(x, y, z)$ found from measured intensities is depicted by electron-density maps on which the lines connect points with equal values of ρ. Because of the smallness of the structure factor, the hydrogen atoms are not directly seen on electron-density maps. Their position can be found from the curvatures of contours of equal electron density.

The three-dimensional distribution of electron density can be made visible, for example, by superimposing contour maps plotted on transparent plastic sheets. The corresponding pattern for myoglobin is shown in Fig. 5.4. The final result of the investigation is the

three-dimensional model of the protein molecule, in which the positions of all the atoms are located (see pages 126 and 201).

Before the forties of this century the X-ray diffraction patterns of relatively simple low-molecular-weight compounds confirmed their structure established by chemical methods and provided quantitative information on interatomic distances. In 1944 Dorothy Crowfoot Hodgkin deciphered the structure of penicillin which could not be determined by chemists. A molecule of penicillin contains 23 atoms

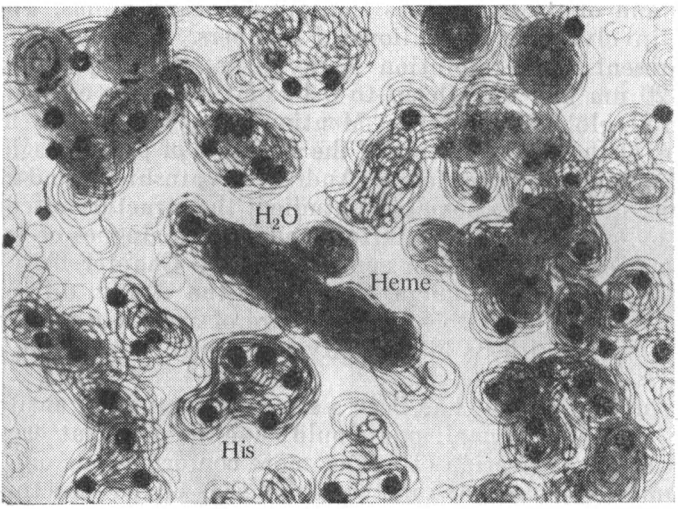


Fig. 5.4. The spatial distribution of electron density in myoglobin.

other than hydrogen atoms. Further, Hodgkin established the structure of vitamin B-12, having determined the coordinates of as many as 93 atoms. Later, the X-ray diffraction technique was employed in investigations of proteins, the most complex molecules. The founder of this trend in molecular biophysics was J.D. Bernal. The greatest achievements in the study of biopolymers are due to the Cambridge research school (Bragg, Kendrew, Perutz). In 1957 John C. Kendrew solved the spatial structure of the first of the proteins analyzed—myoglobin (page 126), having located the positions of its 2500 atoms.

Protein crystals required for X-ray diffraction analysis are very difficult to grow. They are molecular lattices with globular macromolecules residing in their points. It has been found that the denaturation of these molecules (globule-coil transitions) may precede the

melting of the crystal (Esipova and Makarov). Protein crystals contain a large amount of water and they are studied in a mother liquor. Bernal and Hodgkin were the first to use this method; they obtained tens of thousands of clear-cut reflections on X-ray diffraction patterns. Their number may even amount to hundreds of thousands. The interpretation of such complicated X-ray diffraction patterns is a very difficult and time-consuming task that can be undertaken only by using electronic computers. For an accurate determination of a phase corresponding to each reflection it is necessary to measure its intensity several times upon diffraction of a pure protein and of its derivatives containing heavy atoms. The calculations involve tens of millions of numbers.

At present, high-resolution X-ray diffraction (up to 0.25 and even 0.20 nm and for Mb up to 0.14 nm) has been carried out for about 150 globular proteins. Mention should be made here, for example, of the interpretation of the structure of pepsin, leghemoglobin, aspartate aminotransferase (Andreeva, Vainshtein, and Borisov). A number of enzymes have been studied; the structure of complexes formed by enzymes with inhibitors and with analogues of substrates has also been elucidated in some cases (see Chapter 6).

The fundamental problem that faces the X-ray diffraction of proteins is as follows: Does the structure of a protein molecule in a crystal coincide with its biologically functional structure in aqueous solution?

If the protein globule had been a strongly fluctuating entity rather than an aperiodic crystal, one would have thought that crystallization means the selection of one or more conformations out of their large number in solution. But the globule has a fixed structure. Crystalline proteins, as has already been said, contain a large amount of water, and they are studied in mother liquor. The results of the X-ray diffraction studies of a protein crystal are, as a rule, in agreement with the data of optical measurements of the same protein in solution. In particular, the degrees of helicity determined by both methods coincide. Moreover, it has been found that the enzymic activity of a number of proteins is retained in a strongly aqueous crystal. On the whole, the answer to this question is positive (see the next section). On the contrary, proteins that are subjected to lyophilic drying change their structure—the X-ray diffraction patterns of dried protein crystals are highly impoverished in reflections.

An especially accurate and delicate procedure has recently been used to study the temperature dependence of the diffraction of X-rays by crystalline proteins (metmyoglobin, myoglobin, lysozyme) and to determine the mobility of the amino-acid residues. This is important for the physics of enzymes (see Sec. 6.4).

The X-ray diffraction patterns of fibrous biopolymers (collagen, DNA) that form systems which are periodic in one dimension are

characterized by some specific features. Figure 5.5 shows the X-ray diffraction pattern of the sodium salt of DNA in the so-called *A*-form (see page 239). A fibrous system is characterized by the extension of diffraction spots into so-called *layer lines* and their predominant disposition near the meridian or equator of the diffraction pattern. The observed picture is the so-called *axial texture* of a unidimensionally ordered system in which some axes of the molecule are nearly parallel to one chosen direction and the other axes are oriented randomly. The texture has one axis of symmetry of infinite order.

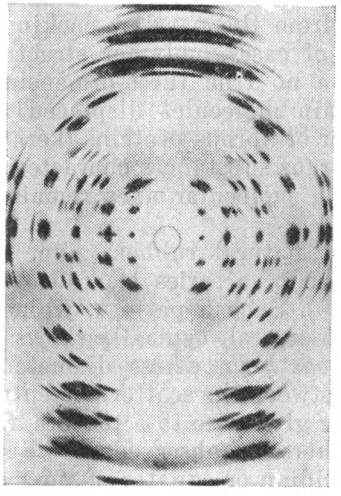


Fig. 5.5. The X-ray pattern of the Na-salt of DNA in a crystalline A-form.

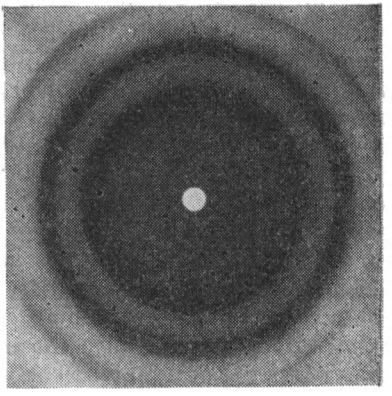


Fig. 5.6. The Debye photograph of collagen.

A number of biopolymers, say DNA, characteristically have a helical structure. In this case, the diffraction pattern of the texture shows a predominant location of diffraction spots along two straight lines, which form an oblique (Andrew) cross (Fig. 5.5). One should not think, however, that a simple observation of an X-ray diffraction pattern and the presence of an oblique cross are sufficient for the interpretation of the helical structure. The object under study is an aggregate of chain molecules which is much less ordered than a three-dimensional crystal. Accordingly, the diffraction pattern is impoverished in reflections. A number of techniques have been developed for the interpretation of diffraction patterns. For example, the trial-

and-error method is employed: a trial model of the system is constructed on the basis of structural-physical concepts and analysis of atomic models, and the distribution of the reflection intensities is calculated. If this distribution coincides with the observed distribution, this proves the correctness of the model. Further, it is possible to calculate the phases and derive the electron-density distribution of the biopolymer. It is by means of this procedure that the famous double helix of DNA was established (see Sec. 7.2).

Aggregates of chain molecules exist in various degrees of orderliness, from a true crystal to an amorphous polymer, in which the chains are disordered. A rigorous theory allows one to judge, from the diffraction pattern, about the disturbances of order caused by displacements, bends, and departures from the parallel packing of macromolecular chains. In a number of cases, polymer structures form *paracrystals*—systems which have no true three-dimensional order but which are composed of chain molecules displaced and turned parallel to each other. A number of fibrous proteins (keratin, collagen), cellulose and some other fibrous substances characteristically exhibit disturbances of order with molecular axes remaining nearly parallel to each other.

Synthetic amorphous polymers (say, natural rubber) give, like liquids, diffraction patterns in the form of a series of concentric rings (*Debye-Scherrer diagrams*, Fig. 5.6). Such a pattern, which is much poorer than the Laue diagram of a crystal, typically shows the presence of a diffuse ring, an *amorphous halo*, whose diameter is determined by predominant distances between the scattering centres. The extension of an amorphous polymer gives rise to a texture and, instead of rings of uniform intensity, there are observed, as we see, more or less extended arcs near the meridian or equator of the ring. Similar patterns are shown by fibrous proteins and also by supermolecular structures of the type of muscle fibres.

5.2. Diffuse Scattering of X-Rays by Solutions of Biopolymers

The diffuse scattering of X-rays by solutions of macromolecules yields direct information on the distribution of scattering objects. Diffuse scattering is the total scattering by randomly arranged individual macromolecules, for which reason the scattered light intensities are averaged over all possible orientations of the macromolecule. The larger the scattering angles, the smaller are the sizes of structural details on which X-rays are diffracted. Small-angle scattering is used to study the gross structure of macromolecules in solution; the specific features of their internal structure are studied at medium and large scattering angles.

The theory of small-angle diffraction is based on conceptions similar to those employed in the theory of light scattering by solu-

tions of macromolecules (see page 89). The theory enables one to relate the scattering intensity observed at the various angles, i.e., the scattering indicatrix, to the distances between scattering particles. In order to determine the shape of the macromolecule, one has to make use of some assumptions—the macromolecule is considered to have the shape of a sphere, ellipsoid or extended cylinder. For these and also for other simple bodies one calculates the scattering indicatrix as a function of the geometric parameters of the macromolecule. For example, for a sphere one calculates the electronic

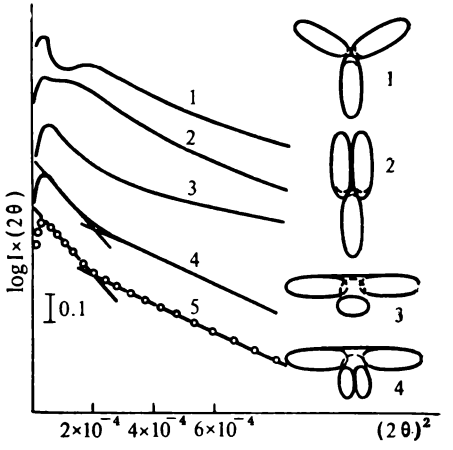


Fig. 5.7. Theoretical scattering indicatrices for various models of immunoglobin (1-4) and experimental points (5).

radius of gyration (since X-rays are scattered by electrons). For myoglobin the radius of gyration has been found to be equal to 1.6 nm, which is in agreement with the dimensions determined by the X-ray structure analysis of crystalline myoglobin. If the scattering system is stretched, the electronic radius of gyration of its cross-section is determined. Scattering indicatrices have been used for the determination of the size, shape, and molecular weights of a number of biopolymers. For example, lysozyme is represented by an equivalent ellipsoid of rotation with dimensions of $2.8 \times 2.8 \times 5.0$ nm^3. More detailed information on the shape of homogeneous particles is obtained by analysis of scattering curves at large angles (from several degrees to about 25° for the CuK_α radiation). Large-angle scattering characterizes structural details with dimensions from 4 to 0.5 nm.

Figure 5.7 shows the theoretical scattering indicatrices calculated for various models of γG_1-immunoglobin and the experimental points lying on the theoretical curve that corresponds to the T-shaped

model of the arrangement of domains (see page 133). This model was later fully confirmed by X-ray crystallographic investigations.

In calculations of X-ray scattering it is necessary to take account of the effect of solvents on the shape of the indicatrix. This effect is not specific and the solvent may be replaced by a homogeneous and structureless continuum characterized by only one parameter—the effective electron density.

Ptitsyn assumed that native proteins display certain mobility of the domain structure (page 120). X-ray diffraction analysis does also shows limited mobility in crystals (page 146). Large-angle scattering in solution has also revealed such mobility in some cases. The method consists in comparing the experimental indicatrix with the theoretical one calculated on the basis of crystallographic data. No noticeable differences have been found between the structures in a crystal and in solution (in the range of distances $\geqslant 1$ nm) for ribonuclease and for complexes of ribonuclease and lysozyme with their inhibitors. The experimental curve for lysozyme without an inhibitor, however, differs slightly from the theoretical curve in the range of distances of about 2 nm. This difference may be interpreted as a fluctuating expansion of the gap between two protein domains. Appreciable differences have been noted on the curves for myoglobin, which is evidently accounted for by fluctuating displacements of its GH-hairpin relative to the other parts of the molecule. There have also been established substantial differences in structural data for crystalline and dissolved phage T4 lysozyme and for other proteins. Thus, at least some proteins in solution are characterized by displacements of domains that make up globules.

5.3. Methods of Nuclear Physics

Modern nuclear physics, which makes use of powerful beams of elementary particles—electrons, neutrons, etc.—offers also such methods as X-ray diffraction using *synchrotron radiation* and *neutron diffraction* for crystalline and dissolved biopolymers.

Magnetic bremsstrahlung, or synchrotron radiation, arises during the motion of charged particles, say electrons, in cyclic accelerators. In synchrotron radiation the energy spent for acceleration of particles is lost. Synchrotron radiation in the X-ray region of the spectrum has an intensity two orders of magnitude higher than the modern X-ray tube with a power of 2 kW. The use of special monochromatization and collimation enables a two-order increase in intensity of X-ray synchrotron radiation; a further increase by 2-3 orders is achieved with the aid of special storage devices. Synchrotron radiation is characterized by a small divergence of the beam. These specific features allow one to make use of synchrotron

radiation for the X-ray diffraction of rapidly changing materials.

Synchrotron radiation is a very promising method for the study of biological processes associated with conformational and other structural transformations at the molecular and supermolecular levels.

Synchrotron radiation offers important possibilities for X-ray spectroscopy. X-ray quanta have an energy just enough for the inner electrons in the atom to be ionized. The X-ray absorption edge shows the fine structure of the absorption band. This structure is the result of the reflection of the spherical wave, emitted by the primary absorber (say, a heavy atom located at the centre of the molecule), from the surrounding atoms and of the interference with the waves scattered by them. The use of this method, called EXAFS (the extended X-ray absorption fine structure technique), has made it possible to obtain valuable information on the structure of *nitrogenase*—a protein containing molybdenum atoms.

Microparticles undergo diffraction if the de Broglie wavelength defined by

$$\lambda = \frac{h}{mv} \tag{5.2}$$

where h is Planck's constant, m is the mass of the particle, and v is its velocity, has the same order of magnitude as the distance between the details of the object. Neutron diffraction has found various applications.

A beam of neutrons from an atomic reactor is subjected to monochromatization by reflection from a crystalline plate (say, CaF_2). Diffracting neutrons are recorded by counters. The neutrons are scattered not by the electronic shell of the atom but by its nucleus, and the atomic scattering factor is determined by the proton-neutron structure of the nucleus rather than by the atomic number. Therefore, the atomic factors of isotopes vary significantly. The atomic factor for hydrogen, i.e., for the proton, is far from being minimal. Therefore, in contrast to X-ray diffraction, neutron diffraction allows hydrogen atoms to be reliably located. For instance, neutron diffraction has been employed for establishing the detailed structure of ice; it has been shown that each hydrogen atom, which forms covalent and hydrogen bonds

$$O—H \cdots O$$

may occupy two positions—the one shown above and the following one:

$$O \cdots H—O$$

In other words, the proton may reside near its own oxygen atom and near a foreign oxygen atom. The hydrogen bond in water is charac-

terized by the jumps of protons between two equivalent minima. Neutron diffraction has been successfully employed in investigations of biopolymers. As regards the diffuse scattering of neutrons by solutions of biopolymers, it yields results which substantially supplement those obtained by scattering of X-rays. By changing the fraction of heavy water in an aqueous solution of a biological object one can vary the overall amplitude of neutron scattering. This reveals the selective scattering of different functional groups. It has been possible, for example, to investigate the nature of the complex formation of a protein and DNA in chromatin.

A specific method of structural studies of biopolymers is *γ-resonance spectroscopy* based on the *Mössbauer effect*. This effect consists of the resonance absorption by an atomic nucleus of the monochromatic γ-radiation emitted by a radioactive atom. All natural compounds of iron contain 2.2 percent of the isotope ^{57}Fe. This makes it possible to study Fe-containing proteins (hemoglobin, myoglobin, cytochromes, etc.) by the γ-resonance method. The sample is exposed to γ-rays emitted by radioactive ^{57}Fe (produced from ^{57}Co). If the source is made to move at a low velocity relative to the absorber, i.e., the sample under study, along the axis that connects the source and the sample, the frequency of the γ-quantum will be changed a little due to the Doppler effect. At a certain value of natural frequency the ^{57}Fe nuclei of the sample will absorb γ-quanta resonantly, passing from the ground state to an excited state. The resonance frequency of the ^{57}Fe nucleus in heme and other Fe-containing groups is highly sensitive to the action of neighbouring atoms and groups. This enables one to obtain information on the electronic structure of Fe-containing atomic groups in their various states in proteins. The effect can be enhanced by enriching a biopolymer with the ^{57}Fe isotope. For this purpose, the content of this isotope is increased in the food of the animals under test, which are the source of the appropriate substances. Apart from heme-containing proteins, the method of γ-resonance spectroscopy is used effectively to study iron-sulphur proteins (ferredoxin, rubredoxin, etc.) and also the mechanisms of chemical reactions involving iron-containing biopolymers. For structure determinations one can use γ-emitting isotopes as labeled atoms and the Mössbauer effect as the analyzer.

Apart from absorption γ-resonance spectra, emission spectra are also studied. In this case, the source contains atomic nuclei that are parent nuclei with respect to Mössbauer nuclei. The emission method has been used for studying the structure of a number of porphyrin compounds of cobalt—vitamin B_{12} and others. The sensitivity of the emission method is very high; this method can be used to follow fast biological processes.

The prospects of the application of γ-resonance spectroscopy are associated, first, with the development of Mössbauerography, i.e.,

the X-ray diffraction based on the radiation of Mössbauer nuclei. Second, the development of γ-lasers has begun, which will make it possible to use holography at a molecular level.

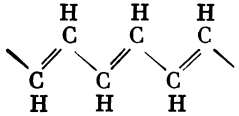
•

5.4. Electronic Absorption Spectra

In the long run, all optical and spectroscopic properties of molecules, which yield important information on their structure, are determined by the distances between the energy levels of the molecule and by the probabilities of transitions between them. These levels correspond to various electronic, vibrational, and rotational states of molecules.

The electronic transitions of polypeptide and polynucleotide chains and, hence, of proteins and nucleic acids occur in the ultraviolet region of the spectrum. The absorption bands of the peptide linkage —CO—NH— lie in the region of 185-240 nm; the same region and a shorter-wave region contain the absorption bands of aliphatic side chains of amino-acid residues. The aromatic residues Try, Phe, and Tyr possess absorption bands in the 280 nm region. The nitrogenous bases in nucleic acids absorb light in the 260-nm region. Thus, proteins and nucleic acids are colourless; they do not absorb visible light.

This is in full agreement with the dielectric nature of biopolymers and renders the statements of their electronic conduction less likely. We have already said that for electrons to be freely transferred along the polymeric chain, the latter must be a rigid conjugated system consisting of alternating double and single bonds or of conjugated aromatic rings. The chain flexibility is incompatible with its electronic conductivity. Short conjugated chains are coloured (carotene is red; see page 52), and sufficiently long chains absorb all the visible light, i.e., they are black and even metal-like. For example, the following polymer of acetylene HC≡CH has been recently obtained:



which is black and has a metallic lustre. Quantum-mechanical calculations show that for conduction electrons in proteins to be excited a very large amount of energy is required, not less than 3.5 eV (i.e., 340 kJ/mole).

Some crystalline semiconductors are colourless, but this does not imply at all that biopolymers may be considered to be semiconductors. Evidently, the low electronic conductivity of biopolymers is accounted for by the presence of metal ions which are difficult to remove and is of no biological significance.

The monomer of the protein chain is similar to the *amide* molecule $R_1—CO—NH—R_2$. As a result of the study of the electronic spectra of simple amides (formamide, methylacetamide, myristamide), there have been established electronic transitions in the amide group (Fig. 5.8). The scheme of electronic transitions is based on quantum-mechanical calculations. Figure 5.9 shows the wavefunctions of the amide group corresponding to its most mobile electrons. The π_2 level in Fig. 5.8 corresponds to the bonding orbital and the π^* level to the nonbonding orbital of the C=O bond; the π_1 level corresponds

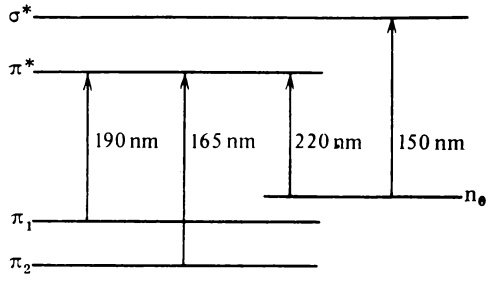


Fig. 5.8. Electron transitions in the amide group.

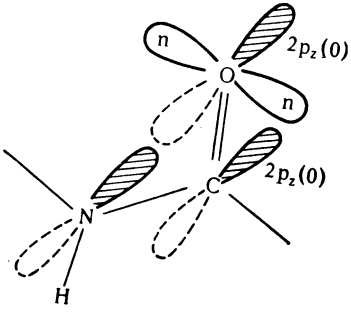


Fig. 5.9. The wavefunctions of the amide group.

to the nonbonding orbital of nitrogen and the n level to the state of the unshared electron pair of oxygen. The $n \rightarrow \sigma^*$ transition is an atomic electronic transition in oxygen.

According to the *Beer-Lambert law*, the intensity of light that has passed through a layer of an absorbing substance of thickness l is equal to

$$I = I_0 \exp(-\varepsilon c l) \tag{5.3}$$

where I_0 is the incident intensity, c is the concentration of the absorbing substance (in moles/litre), ε is the molar extinction coefficient (or molar absorptivity) [in litres/(mole·cm)]. The absorption in a given band, which corresponds to the transition from the ground level 0 to an excited level j is expressed in terms of the so-called *oscillator strength* f_{0j}, which is determined by the intensity of the absorption band:

$$f_{0j} = 4.32 \times 10^{-9} \int \varepsilon_{0j}\, d\tilde{\nu} \tag{5.4}$$

Here $\tilde{\nu}$ is the wave number, i.e., the frequency of the light wave divided by the light velocity; the integration is carried out over the

entire absorption band. Theory yields

$$f_{0j} = \frac{2m_e}{3\hbar e^2}\,\omega_{0j} p_{0j}^2 \tag{5.5}$$

where e and m_e are the charge and mass of an electron; ω_{0j} is the circular frequency corresponding to the $0 \rightarrow j$ transition:

$$\omega_{0j} = (E_j - E_0)/\hbar \tag{5.6}$$

The quantities E_0 and E_j are the energies of the ground and excited states. Finally, p_{0j} is the matrix element of the dipole moment corresponding to the $0 \rightarrow j$ transition or the *transition dipole moment*:

$$p_{0j} = \int \psi_0 p \psi_j^* \, d\tau \tag{5.7}$$

The quantity p_{0j}^2 is called the *dipole strength* of the transition. The oscillator strength is a dimensionless quantity, whose values for different bands vary from very small to those of the order of 0.1-1.0. Formula (5.7) indicates that p_{0j}^2 and the oscillator strength are the greater, the more strongly the wavefunctions of the initial and final states, ψ_0 and ψ_j, overlap. The $n \rightarrow \pi^*$ transition in the amide group has a low dipole strength since the electron clouds of the n- and π^*-states are almost perpendicular to each other and overlap only slightly (Fig. 5.9).

Theoretical calculations give the following values of wavelengths λ_{0j} which correspond to the following transitions: $n \rightarrow \pi^*$ at 234 nm, $\pi_1 \rightarrow \pi^*$ at 160 nm and $\pi_2 \rightarrow \pi^*$ at 135 nm. The spectrum of myristamide shows bands of 220 nm ($n \rightarrow \pi^*$), 190 nm ($\pi_1 \rightarrow \pi^*$), and 165 nm ($\pi_2 \rightarrow \pi^*$). The transition dipole moment of the $\pi_1 \rightarrow \pi^*$ transition lies in the plane of HNCO at an angle of 9° to the line connecting the atoms O and N, and the transition moment of the $n \rightarrow \pi^*$ transition is perpendicular to this plane.

The spectra of amide groups in a polymeric chain vary because of the electronic interaction between these groups. For example, the energy levels for the terminal groups of the chain are different from those for the internal groups. Of special interest in biophysics are the spectral effects that arise as a result of the exciton resonance interaction, namely the *Davydov splitting* and *hypochromism*.

Studying the optical properties of molecular crystals, Davydov showed that resonance transfer of the excitation energy may occur between the excited energy levels of a regular set of identical chromophoric (light-absorbing) groups. The excitation energy, or *exciton*, can propagate in a regular system (Frenkel). As a result of the resonance interaction, the energy levels are split, producing a wide zone. The selection rules allow transitions on absorption and emission of light not to any but only to strictly definite levels of the zone, the

polarization of absorption bands being determined by the symmetry of the regular system as a whole.

A resonance interaction is possible even in a dimer, in which it can lead to the splitting of the ground level into two levels. Figure 5.10 shows the scheme of such splitting for the cases of the parallel and collinear arrangement of transition dipoles. The solid arrows indicate allowed transitions, and the broken arrows show forbidden transitions. Moffitt has carried out a theoretical calculation of the

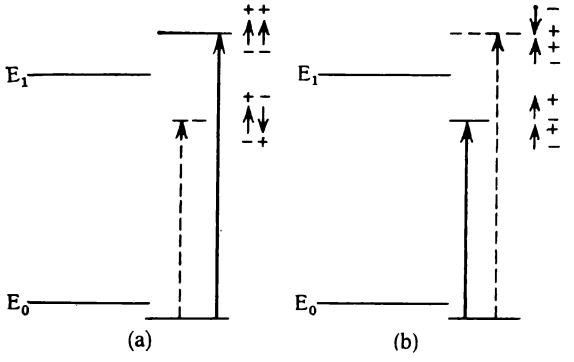


Fig. 5.10. Energy levels for a system of parallel (*a*) and collinear (*b*) transition dipole moments.

exciton splitting for the $\pi_1 \rightarrow \pi^*$ transition at 190 nm in the α-helix. According to the Davydov theory, this band splits into two bands, one of which is polarized along the helix axis and the other at right angles to it. The frequency difference between the two bands is 2700 cm^{-1} (or about 10 nm on the wavelength scale).

Resonance interaction leads to a redistribution of the spectral line intensities. In the case of two collinear transition dipoles the band of lower frequency (greater wavelength) shows an increased intensity at the expense of the intensity of the short-wavelength band. The long-wavelength band shows a hyperchromic effect. On the contrary, in the case of parallel dipole moments the intensity of the long-wavelength band falls off and that of the short-wavelength band is increased. Now the long-wavelength absorption band displays a hypochromic effect. It is exactly hypochromism that is observed in the spectra of α-helical polypeptides and proteins and also in double-stranded helical nucleic acids. If the dipole moments are perpendicular to each other, no intensity redistribution will occur.

Figure 5.11 shows the experimental results obtained for polyglutamic acid (PGA). At pH 4.9 the polymer has an α-helical form and at

pH 8.0 this changes to a coil. Two bands are clearly seen, which have arisen as a result of a Davydov splitting, and a strong hypochromism is observed.

The oscillator strength of the long-wavelength band of the α-helix is decreased by 30 percent. The long-wavelength bands of polyglutamic acid and polylysine being in a coiled form are characterized by an oscillator strength of 0.110; the oscillator strength for the α-helix is 0.075.

The hypochromic effect has been examined in detail both experimentally and theoretically. The dependence of the effect on the medium, pH, and ionic strength has been studied.

It is obvious that the disappearance of hypochromism in the course of a helix-coil transition can yield a quantitative measure of the α-helicity of proteins. Because of the difficulties involved in spectrophotometric measurements in the far-ultraviolet region near 200 nm, this method is used seldom in studies of proteins. On the contrary, it is simple and efficient for nucleic acids, whose long-wavelength bands lie near 260 nm.

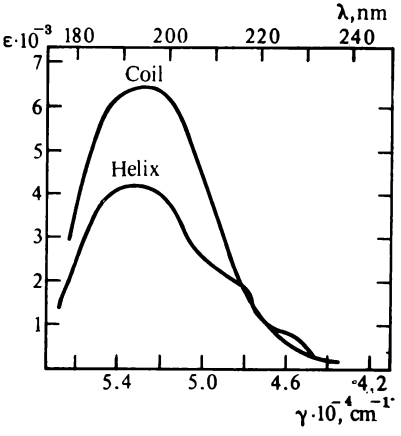


Fig. 5.11. The absorption spectrum of polyglutamic acid (PGA).

5.5. Luminescence

Like absorption spectra, the luminescence (fluorescence) spectra of complex molecules are blurred and devoid of fine details. This does not refer to *Shpolsky line spectra*, which are observed under special conditions and which have so far found no application in biophysics. The intensity, polarization, and duration of fluorescence prove to be more informative than the wavelengths of band maxima. The wavelengths of band maxima are not so much informative as the intensity, polarization, and duration of fluorescence.

Suppose that at time zero ($t = 0$) there are n (0) excited molecules. If the probability of the transition for 1 s from an excited to an unexcited state supplemented by the emission of a light quantum is f and the probability of a radiationless transition, in which the excitation energy is converted to heat, is g, then

$$dn = -(f + g)\, n\, dt \qquad (5.8)$$

and

$$n(t) = n(0) \exp[-(f+g)t] \tag{5.9}$$

The average lifetime of the excited state is given by

$$\tau = \int_{t=0}^{\infty} t\,dn \Big/ \int_{t=0}^{\infty} dn = \frac{1}{f+g} \tag{5.10}$$

The intensity of fluorescence decreases exponentially. In most cases τ is of the order of magnitude equal to 10^{-8}-10^{-9} s, but it may be much greater.

The quantum yield is expressed in terms of fractions of radiative transitions:

$$\gamma = f/(f+g) \tag{5.11}$$

In solution the equilibrium distribution of molecules over their reserves of vibrational energy is independent of the excess of this energy produced on excitation and, hence, of the wavelength of the exciting light, λ_{exc}. Thus, τ and γ are also independent of λ_{exc} (the *Vavilov law*).

The fluorescence emission of complex molecules (dyes, in particular) is polarized even with a natural incident light. The theory of luminescence polarization has been worked out by Vavilov and Feofilov. The exciting light is absorbed by molecules that are oriented in a special manner with respect to the electric vector of the light wave. After the light has been absorbed the energy is emitted as a result of another electronic transition, to which there corresponds, generally speaking, a different polarization in the molecule, i.e., a different direction of the transition dipole. If the lifetime of the excited state, i.e., the energy transfer time, is short as compared with the reorientation time of the molecule, the luminescence becomes polarized. The *degree of polarization* is defined as follows:

$$P = \frac{I_z - I_x}{I_z + I_x} \tag{5.12}$$

where I_z and I_x are the radiation intensities with the components of the electric vector along the axes z and x, the incident light being directed along the y axis (Fig. 5.12). Calculations show that the maximum value of P with a natural incident light for fixed, randomly oriented molecules is 1/3. The rotational motion of molecules or atomic groups depolarizes the fluorescence. Theory yields the *Levshin-Perrin formula*

$$\frac{1}{P} = \frac{1}{P_0} + \left(\frac{1}{P_0} + \frac{1}{3}\right)\frac{kT}{V\eta}\tau \tag{5.13}$$

Here P is the observed degree of polarization; P_0 is its maximum value in the absence of depolarization, i.e., at $T \to 0$, or at the vis-

cosity of the medium, η, tending to infinity; V is the volume of the molecule; τ is the lifetime of the excited state. The degree of polarization falls off with increasing mobility of the light-emitting molecule and may serve as a measure of mobility.

Radiationless transitions, which are responsible for the quantum yield γ being less than unity, reduce to the conversion of light energy into heat. The molecule excited by a photon $\hbar\omega_a$ to a certain singlet

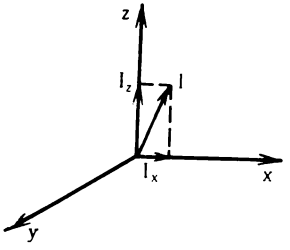


Fig. 5.12. Diagram illustrating luminescence polarization.

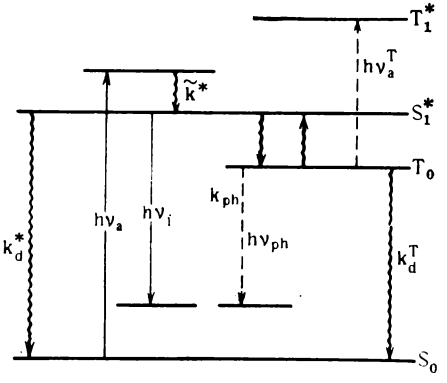


Fig. 5.13. Energy transfer diagram for luminescence.

level E_{S_1} is capable of emitting a quantum $\hbar\omega_f$ with probability f; the energy can be converted within the molecule into vibrations with subsequent degradation into heat; the molecule can pass, without emission, to a metastable triplet state level E_{T0}, expending part of the excitation energy. The energy E_{T0} can then be released as a phosphorescence photon $\hbar\omega_{ph}$ or else can degrade into heat. All these processes are depicted schematically in Fig. 5.13.

Another important factor responsible for the decrease in quantum yield is the *quenching of luminescence*. It can be provoked by foreign substances or by the migration of the excitation energy from molecule to molecule (concentration quenching). We are not speaking here of the emission with subsequent reabsorption of light but of a direct transfer of excitation energy due to the resonance interaction of molecules which increases with decreasing distance between molecules, i.e., with increasing concentration. The condition for such an interaction is a strong overlap of the absorption and emission bands of the molecules.

The difference between the resonance and exciton (page 155) interactions is of a quantitative nature. The resonance migration of

energy occurs if the interaction energy is great, i.e., by far exceeds the width of the electronic-vibrational zone. In this case, energy transfer takes place rapidly and the effect of vibrations may be neglected. In contrast to exciton spectra, which arise in the case of a weak interaction (the energy of which is much less than the width of the zone), in the spectra of resonantly interacting molecules no vibrational structure is observed.

The migration of energy brings about the depolarization of luminescence. A molecule that absorbs a quantum is oriented in some way in space. If the same molecule emitted light, it would have a certain value of the degree of polarization P. However, during the time τ between absorption and emission there can repeatedly occur acts of energy transfer to other molecules with somewhat different orientations. After a certain time period has elapsed all the excited molecules lose their original orientation and the emission becomes depolarized.

It should be emphasized that energy migration is possible between different as well as identical molecules provided that the absorption bands of molecules of one species and the fluorescence bands of molecules of another species overlap.

The study of the specific features of luminescence associated with migration of energy is important for biophysics. The transfer of energy from one luminescent part of a biopolymer to another allows one to estimate their relative arrangement. The study of polarized luminescence yields information on the conformational structure and its dynamics—on the mobility of the macromolecule and of its individual groups. The frequently used luminophors are cyclic π-electronic systems. For example, aromatic amino-acid residues in proteins emit luminescent radiation. The spectral properties of the corresponding amino acids are presented in Table 5.1.

Table 5.1. The Spectral Properties of Aromatic Amino Acids

Amino acid	λ_{max}, nm (absorption)	ε, mole^{-1}·litre^{-1}	λ_{max}, nm (fluorescence)	γ
Phenylalanine	257	200	282	0.04
Tyrosine	257	1300	303-304	0.21
Tryptophan	280	5600	353	0.21

The fluorescence of Phe is not practically observed. The luminescence of Tyr and Trp depends substantially on the pH and polarity of the medium and on the protonation of neighbouring groups. In native proteins, the value of γ for Typ is much less than 0.21; on the contrary, the value of γ for Trp may increase up to 0.32. Denaturation with urea makes the values of γ for Tyr and Trp equal—their quantum yields are conformationally sensitive. The point is that

the quenching of the luminescence of Tyr and Trp depends on the environment, particularly on its hydrophobic nature. The luminescence of residues that are transferred during conformational changes from the interior of the globule to its surface is quenched more easily by foreign quenchers. The study of the luminescence of proteins (aromatic residues) has furnished valuable information about their conformational properties and interactions with other molecules and ions.

No less valuable results are obtained by the application of luminescent labels. The conformational structure and dynamics of complexes of nucleic acids with luminophors (acrydine dyes) are successfully studied with the aid of polarized luminescence.

Three trends in biophysical investigations are directly or indirectly associated with luminescence. Some critical remarks concerning these trends are in order.

1. The concepts of energy migration which were developed through the study of luminescence formed the basis for attempts to explain a number of biological processes. Many of these processes occur rapidly, according to the "all-or-none" principle. A true explanation of such phenomena must presumably be based on their cooperative nature. Such an explanation has been given for many problems. In the literature, however, there are certain considerations as to the migration of energy in a biopolymer or in the seemingly "structured" water that surrounds it. These considerations were used to explain, for example, muscular contraction. The "migration" conceptions in the "dark" biology have no theoretical backing.

2. In 1923 Gurvich reported the discovery of *mitogenetic rays* which are emitted by cells during their mitotic division and which in their turn stimulate mitoses. The attempts made for a number of years to detect this emission by means of precise physical methods were not successful. The existence of mitogenetic rays has not been verified and therefore their study has long been abandoned. Other reports concerning specific ultraviolet radiations that arise in other biological processes have likewise proved unsubstantiated.

3. The cells and tissues of plants and animals often produce a weak fluorescence in air which was treated as being of biological nature. This luminescence, however, is in fact *chemiluminescence*, which is mainly determined by the oxidation of lipids. Many organic compounds fluoresce on oxidation. Apparently, this fluorescence is of no biological value. It can be used as an indicator of oxidation processes.

5.6. Optical Activity

As we have already said, biopolymers are chiral and, hence, display natural optical activity, i.e., rotate the plane of polarization of plane-polarized light, and *circular dichroism* (which will be discussed later).

Optical activity, the ability of a chiral medium to rotate the plane of polarization of transmitted light, was discovered by Arago in 1811. As has been shown by Fresnel (1820), optical activity is the result of *circular birefringence*, i.e., of different velocities of propagation in a medium of the right and left circularly polarized light. In the right-hand wave the electric field strength vector in the ray directed towards the observer rotates in a clockwise direction; in the left-hand wave the vector rotates in an anticlockwise direction. A linearly polarized wave can be resolved into a right-hand and a left-hand circularly polarized components—the electric vector oscillating along a single direction is the sum of two vectors which rotate circularly in a clockwise and an anticlockwise direction (Fig. 5.14*a*). If one of these components propagates faster than the other, the resultant vector rotates at an angle that is the larger, the greater is the velocity difference (Fig. 5.14*b*), i.e., the higher is the difference of the refractive indices. When the ray is passed through a layer of a substance of thickness l, the plane of polarization is rotated through an angle given by

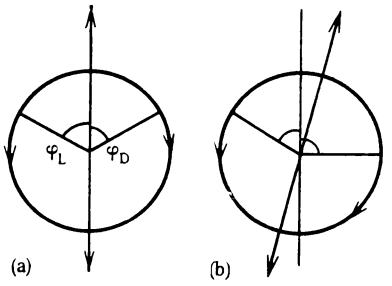


Fig. 5.14. Resolution of a plane-polarized wave into right- and left-circularly polarized components (*a*) and rotation of the plane of polarization caused by circular birefringence (*b*).

$$\varphi = \frac{\pi}{\lambda}(n_L - n_D)\, l \tag{5.14}$$

where λ is the wavelength of light; n_L and n_D are the refractive indices for the left-hand and right-hand waves.

The specific rotation of a solution containing c (g/cm^3) of an optically active substance is defined by the following relation:

$$\varphi_{sp} = \varphi / lc \tag{5.15}$$

If the thickness l is measured in dm and φ in radians, then

$$\varphi_{sp} = \frac{180}{\pi}\,\frac{10}{c}\,\frac{\varphi}{l}\ \text{rad}\cdot\text{dm}^{-1}\cdot\text{g}^{-1}\cdot\text{cm}^3 \tag{5.16}$$

For a pure substance c is replaced with density. The *molar rotation* is a quantity defined by

$$\mathscr{M} = \frac{M}{100}\varphi_{sp} = \varphi\,\frac{18}{\pi}\,\frac{M}{cl} \tag{5.17}$$

where M is the molecular weight of the optically active substance.

Any substance both refracts and absorbs light. In the region of absorption of light by a substance the refractive index is a complex quantity defined by the formula

$$\tilde{n} = n - i\varkappa \tag{5.18}$$

where $\varkappa$ is the absorption coefficient related to the molar extinction coefficient ε (see formula 5.3) by the following formula

$$\varkappa = \varepsilon \frac{\lambda c}{4\pi} \frac{1000}{M} \tag{5.19}$$

where M is the molecular mass and c is expressed in g/cm^3. Use is often made of the Beer-Lambert law in the form

$$I = I_0 \cdot 10^{-\varepsilon' c' l} \tag{5.2}$$

Here c' is expressed in moles·litre^{-1} and $\varepsilon' = 0.4343\varepsilon$.

If a substance exhibits circular birefringence, it absorbs left- and right-circularly polarized light differently, i.e., it displays circular dichroism as well. The angle of rotation of the plane of polarization is also a complex quantity:

$$\tilde{\varphi} = \frac{\pi}{\lambda} (\tilde{n}_L - \tilde{n}_D) l = \frac{\pi}{\lambda} (n_L - n_D) l - i \frac{\pi}{\lambda} (\varkappa_L - \varkappa_D) l \tag{5.21}$$

In the absorption range, linearly polarized light is transformed into elliptically polarized light. A measure of ellipticity is the quantity

$$\theta = \frac{\pi}{\lambda} (\varkappa_L - \varkappa_D) \tag{5.22}$$

Molecular ellipticity is defined as follows:

$$\theta_{mol} = \theta \frac{18}{\pi} \frac{M}{c} = 2.303 \times \frac{4500}{\pi} (\varepsilon'_L - \varepsilon'_D) \equiv 3300 (\varepsilon'_L - \varepsilon'_D) \tag{5.23}$$

Theory must account for the origin of the differences $n_L - n_D$ and $\varkappa_L - \varkappa_D$ (i.e., $\varepsilon_L - \varepsilon_D$) and establish the relationship between these quantities and the structure of the substance. In conventional optics—the theory of light refraction and dispersion—the sizes of molecules are considered to be infinitesimally small as compared with the light wavelength λ. In other words, no account is taken of the differences in the phases of the light wave at different points of the molecule. In fact, for small molecules the ratio of their sizes to the wavelength of visible light is of the order of 10^{-3}, which cannot be ignored in the theory of optical activity—the entire effect is determined precisely by the phase differences of the light wave within the molecule. Spectropolarimetry is a kind of *molecular interferometry*, which is what accounts for the high sensitivity of optical activity towards changes in molecular structure.

Thus, one cannot neglect the distances between the electrons of the molecule which are acted on by the light wave. It is also necessary to take account of the interaction between these electrons—in the absence of such an interaction no optical activity is present. And, finally, the molecule or crystal must be chiral, i.e., they must have neither plane nor centre of symmetry.

In the rigorous quantum-mechanical theory of optical activity (Rosenfeld) the rotation of the plane of polarization is expressed in terms of electric and magnetic transition dipoles (cf. page 155). For an isotropic medium that contains N_1 of chiral molecules in 1 cm^3 we have

$$\varphi = \frac{16\pi^3 N_1 \beta}{\lambda^2} \frac{n^2+2}{3} l \tag{5.24}$$

where n is the refractive index and β is a constant which characterizes optical activity. The refractive index is related to the polarizability a by the *Lorentz-Lorenz formula*:

$$\frac{n^2-1}{n^2+2} = \frac{4\pi}{3} N_1 a \tag{5.25}$$

The polarizability a, which characterizes the refraction of light, and the constant β, which determines circular birefringence, are defined in a similar way:

$$a = \frac{2}{3\hbar} \sum_j \frac{\omega_j p_{0j}^2}{\omega_j^2 - \omega^2} \tag{5.26}$$

$$\beta = \frac{2\pi c}{3\hbar} \sum_j \frac{\mathrm{Im}\,\{p_{0j} m_{j0}\}}{\omega_j^2 - \omega^2} \tag{5.27}$$

These formulas are valid for molecules in the ground, electronically nonexcited state 0, far away from absorption bands with frequencies $\omega_j \equiv \omega_{0j}$. Here ω is the circular frequency of the incident light; c is the light velocity; $\hbar = h/2\pi$, where h is Planck's constant; $\boldsymbol{p}_{0j}$ is the electric transition dipole moment (see page 155); and $\boldsymbol{m}_{j0}$ is the *magnetic transition dipole moment*. The symbol Im indicates that β incorporates only the imaginary components of the complex products $\boldsymbol{p}_{0j}\boldsymbol{m}_{j0}$. The summation is carried out over all the $0 \rightarrow j$ transitions. As we have already seen (page 155), the quantity $p_{0j}^2 = D_j$, which is called the dipole strength, is directly connected with the intensity of light absorption. The sum of the dipole strengths over all the $0 \rightarrow j$ transitions is a constant which is proportional to the number of electrons in the molecule. The quantity

$$R_j = \mathrm{Im}\,\{\boldsymbol{p}_{0j}\boldsymbol{m}_{j0}\} \tag{5.28}$$

is called the *rotational strength* of the $0 \rightarrow j$ transition.

In contrast to the sum of dipole strengths, the sum of the rotational strengths over all j's is equal to zero.

This can be easily proved:

$$\sum_j R_j = \sum_j \operatorname{Im}\left\{\int \psi_0 \boldsymbol{p} \psi_j^* \, d\tau \int \psi_j \boldsymbol{m} \psi_0^* \, d\tau\right\} =$$

$$= \operatorname{Im}\left\{\int \psi_0 (\boldsymbol{pm}) \psi_0^* \, d\tau - \int \psi_0 \boldsymbol{p} \psi_0^* \, d\tau \int \psi_0 \boldsymbol{m} \psi_0^* \, d\tau\right\} = 0 \qquad (5.29)$$

since the diagonal matrix elements $\langle 0 \mid \boldsymbol{pm} \mid 0 \rangle$, $\langle 0 \mid \boldsymbol{p} \mid 0 \rangle$, and $\langle 0 \mid \boldsymbol{m} \mid 0 \rangle$ are real, i.e., their imaginary parts are equal to zero.

The rotational strength R_j vanishes if the molecule has a plane or centre of symmetry. The electric and magnetic terms cannot be both simultaneously different from zero in the absence of chirality. In such cases all the states of the molecule are specified as even or odd, depending on whether the wavefunction retains or changes its sign upon reflection at the centre or in the plane, respectively. The moment operators $\boldsymbol{p}$ and $\boldsymbol{m}$ in the Cartesian coordinates have the form

$$\boldsymbol{p} = e\,(\boldsymbol{i}x + \boldsymbol{j}y + \boldsymbol{k}z)$$

$$\boldsymbol{m} = -i\frac{e\hbar}{2m_e c}\left\{\boldsymbol{i}\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right) + \boldsymbol{j}\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right) + \right.$$

$$\left. + \boldsymbol{k}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\right\} \qquad (5.30)$$

where e and m_e are the electron charge and mass; $\hbar = h/2\pi$, where h is Planck's constant; c is the velocity of light; $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are the unit vectors along the coordinate axes. Let us assume that the molecule has a centre of symmetry. Upon reflection at the centre $\boldsymbol{p}$ changes its sign and $\boldsymbol{m}$ retains sign. Therefore, $\langle 0 \mid \boldsymbol{p} \mid j \rangle$ is not equal to zero upon transitions from odd to even states and vice versa. On the contrary, $\langle j \mid \boldsymbol{m} \mid 0 \rangle$ is different from zero only for transitions between states of identical parity. Hence, the scalar product of these matrix elements is equal to zero for any transitions. The equality of R_j to zero is proved in an analogous manner for molecules possessing a plane of symmetry.

In going from the right-handed antipode to the left-handed one the sign of R_j, the sign of rotation, is changed. In order of magnitude R_j is equal to the product of the electric and magnetic moments of the molecule. The transition moment $\boldsymbol{p}_{0j}$ is of the order of 1 D $= 10^{-18}$ esu, and the magnetic moment of the electron is 0.93×10^{-20} erg·G^{-1}. Hence, the order of magnitude of R_j is 10^{-38} esu. It is convenient to use the value of the reduced rotational strength expressed by a number of the order of unity:

$$R' = \frac{100}{\mu_B p_d} R$$

where μ_B is the Bohr magneton and $p_d = 1$ D.

In the region of an absorption band of frequency ω_j the damping must be taken into account. Then we obtain a complex expression for β:

$$\beta_j = \frac{2c}{3\hbar}\frac{R_j}{\omega_j^2-\omega^2+i\omega\Gamma_j} = \frac{2c}{3\hbar}\frac{R_j(\omega_j^2-\omega^2)}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2} - \\ - \frac{2ic}{3\hbar}\frac{R_j\omega\Gamma_j}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2} \quad (5.31)$$

The quantity Γ_j characterizes the half-width of the absorption band on a frequency scale. The first term in formula (5.31) describes the

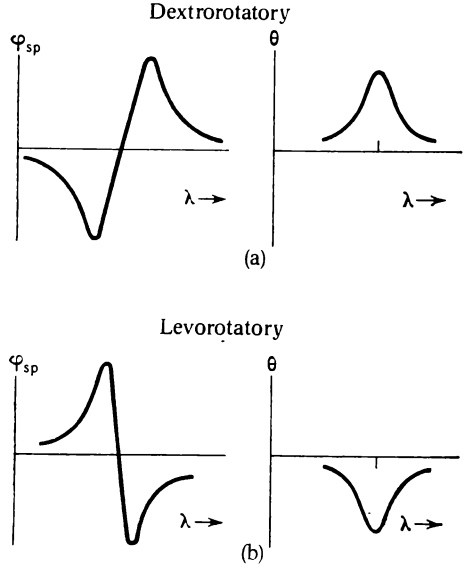


Fig. 5.15. AORD and CD curves for (*a*) a dextrorotatory substance and (*b*) a levorotatory substance.

rotation of the plane of polarization, and the second the circular dichroism (page 163). In an analogous manner we obtain, near the region of absorption, a complex expression for polarizability in which the real term describes refraction and the imaginary one the absorption of light.

Near the region of absorption there are simultaneously observed an anomalous optical rotatory dispersion (AORD) and circular dichroism (CD). The model curves for a levorotatory and a dextrorotatory substance are shown in Fig. 5.15. The corresponding curves for refraction and absorption are analogous to the AORD and CD curves for the dextrorotatory substance. The anomalous optical rotatory

dispersion is usually referred to as the Cotton effect, although Cotton discovered circular dichroism.

The rotational strength R_j can be determined both from AORD and CD. Accordingly, one can pass from one effect to the other through calculation. This is a special case of *Kronig-Kramers relations*, which establish the relationship between the real and imaginary parts of the response of a linear system towards an external action. The calculations of rotation from the ellipticity and vice versa (and analogous calculations for refraction and absorption) are carried out through the use of the formulas

$$\varphi(\omega) = \frac{2\omega^2}{\pi} \int_0^\infty \frac{\theta(\omega')}{\omega'(\omega'^2 - \omega^2)} d\omega' \tag{5.32}$$

$$\theta(\omega) = -\frac{2\omega^3}{\pi} \int_0^\infty \frac{\varphi(\omega')}{\omega'^2(\omega'^2 - \omega^2)} d\omega' \tag{5.33}$$

In practice, the quantity R_j can be conveniently determined in investigations of biopolymers by using CD rather than AORD. It can be shown that the signal-to-noise ratio is much greater in CD.

Theoretical calculations of β are approximate, since it is practically impossible to know the entire set of energy levels of molecules and of the corresponding wavefunctions for direct calculations made by means of formula (5.27).

The molecule can be mentally split into atomic groups with no electrons being exchanged between them. Then the quantity β will be represented by the sum

$$\beta = \beta_1 + \beta_2 + \beta_3 + \beta_4 \tag{5.34}$$

where β_1 are the contributions from individual groups equal to zero if the groups are achiral (say, CH_3 groups); β_2 are the "one-electron" terms which appear as a result of electric and magnetic dipole transitions in a given group, which is in the asymmetric field of the other groups. The term β_3 results from the interaction of $\boldsymbol{m}_{j0}$ of one group with $\boldsymbol{p}_{0j}$ of another. Finally, β_4 is determined by the simultaneous electric dipole excitation of various groups. The contributions of β_2 and β_4 are the most significant. The contribution of β_4 is the result of the electrostatic dipole interaction of anisotropically polarizable groups in the molecule. This contribution is determined by the polarizabilities of atomic groups and depends directly on their mutual geometric arrangement, i.e., on the configuration and conformation of the molecule, and turns to zero in the presence of a plane or centre of symmetry.

The simplest model of a chiral molecule proposed by Kuhn consists of two linear groups, 1 and 2, which are a distance r from each

other and which undergo polarization only along their axes, the angle between which is γ (Fig. 5.16). In this case

$$\beta_4 = \frac{1}{6} a_1 a_2 \sin\gamma \cos\gamma / r^3 \qquad (5.35)$$

where a_1 and a_2 are the polarizabilities of groups 1 and 2. As we have seen (page 164), the polarizability is determined by the values of p_{0j}, i.e., by the intensities of the corresponding absorption bands. The more intense the band, the greater its contribution to β_4. But the rotational strength is given by the product $p_{0j}m_{j0}$. At high values of p_{0j} the values of m_{j0} are usually small, and vice versa. Weak bands with small p_{0j} but large m_{j0} may play a significant role in optical activity. Such bands do not introduce an appreciable contribution to the polarizability; the quantity β_4 takes no account of the role of weak bands.

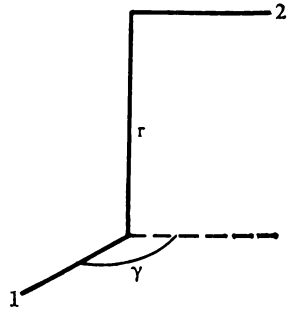


Fig. 5.16. The simplest model of an optically active molecule (after Kuhn).

The contribution of β_2 is estimated approximately with the aid of the so-called *one-electron model.* Formula (5.27) is used but this model is concerned only with electrons of chromophore groups responsible for long-wave absorption. The chromophore group (say, the peptide linkage —NH—CO—) is usually planar and does not exhibit optical activity. But, being in the asymmetric field of the neighbouring atoms of the chiral molecule, the group makes a contribution to β_2. The one-electron model allows one to describe well the optical activity dispersion and, notably AORD and CD, in the absorption band of a chromophore. By using this model it is possible to calculate, with good accuracy, the rotational strengths of electronic transitions in chromophores. A change in conformation of the molecule alters β_2. The theory has been successfully employed for the conformational analysis of terpenes, steroids, etc.

The present-day advancement of spectropolarimetry is directly tied up with investigations into the structure of natural compounds — with bioorganic chemistry and molecular biology. Spectropolarimetry has gone through three stages of development. At one time, molecules and crystals were characterized only by the values of φ_{sp} at a single light wavelength. This gave very scarce information. Later, scientists turned to the measurement of dispersion outside the region of absorption. Finally, the scientists became interested in

measuring the Cotton effect (AORD) and CD in the region of absorption. This is most informative but requires that the apparatus be highly sensitive (that the rotation be measured with an accuracy of up to 10^{-4} deg) and measurements be carried out over a wide range of wavelengths (for proteins, up to 180-230 nm).

5.7. Optical Activity of Biopolymers

For polymers it will be reasonable to express optical rotation and CD in terms of quantities referred to one monomer. The mean rotation of the monomer is defined as

$$m' = \frac{M_0}{100} \varphi_{sp} \tag{5.36}$$

where M_0 is the average molecular weight of the monomer. The effect of the solvent can be approximately taken into account by introducing a Lorentz correction. Instead of m', use is made of the reduced mean rotation of the monomer:

$$\tilde{m'} = \frac{3}{n^2+2} m' = \frac{3}{n^2+2} \frac{M_0}{100} \varphi_{sp} \tag{5.37}$$

where n is the refractive index of the medium. For water at 20°C the Lorentz factor $3/(n^2 + 2)$ varies from 0.7945 at 600 nm to 0.7306 at 185 nm.

In a number of cases, biopolymers exhibit dual chirality—the chirality of monomeric units and the chirality of the α-helix of proteins and polypeptides or of the double-helical portions of nucleic acids. On denaturation, i.e., on helix-coil transitions, only the chirality of monomers is retained. Accordingly, these transitions are accompanied by sharp changes in rotation and CD. These characteristics are very sensitive to changes in conformations. The application of AORD and CD yields valuable information on the conformations of biopolymers and their changes.

Far away from the region of absorption the reduced rotation of the monomer for any molecular system can be expressed in terms of the sum of one-term *Drude formulas* (cf. formula 5.27):

$$\tilde{m'} = \sum_j \frac{a_j \lambda_j^2}{\lambda^2 - \lambda_j^2} \tag{5.38}$$

Here λ is the wavelength of the incident light; λ_j and a_j are constants. Approximately, Eq. (5.38) is replaced with a one-term Drude formula:

$$\tilde{m'} = \frac{a_0 \lambda_0^2}{\lambda^2 - \lambda_0^2} \tag{5.39}$$

This formula describes well the optical rotation of polyamino acids in a coiled state if $\lambda_0 = 268$ nm. For the α-helix one can use the Moffitt formula

$$\tilde{m}' = \frac{a_0^{(\alpha)}\lambda_0^2}{\lambda^2 - \lambda_0^2} + \frac{b_0^{(\alpha)}\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} \tag{5.40}$$

where $a_0^{(\alpha)}$ and $b_0^{(\alpha)}$ are constants. The second term characterizes the contribution of the "second chirality". Formula (5.40) should be considered empirical. The agreement of formula (5.40) with experiment for the α-helix is obtained at $\lambda_0 = 212$ nm and at $b_0^{(\alpha)} = -630$, and $a_0^{(\alpha)} = 630$-650. For a partially helical protein

$$a_0 = a_0^{(aa)} + x_\alpha a_0^{(\alpha)}$$

where $a_0^{(aa)}$ characterizes the rotation due to amino acid residues; x_α is the fraction of helical portions. We obtain

$$\tilde{m}' = a_0^{(aa)} \frac{\lambda_1^2}{\lambda^2 - \lambda_1^2} + x_\alpha \left(a_0^{(\alpha)} \frac{\lambda_0^2}{\lambda^2 - \lambda_0^2} + b_0^{(\alpha)} \frac{\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} \right) \tag{5.41}$$

Expanding the first term into a series in $(\lambda^2 - \lambda_0^2)^{-1}$ and retaining the second term, we obtain

$$\tilde{m}' = a_0^{(aa)} \frac{\lambda_0^2}{\lambda^2 - \lambda_0^2} + b_0^{(aa)} \frac{\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} + x_\alpha \left(a_0^{(\alpha)} \frac{\lambda_0^2}{\lambda^2 - \lambda_0^2} + b_0^{(\alpha)} \frac{\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} \right) \tag{5.42}$$

Usually $b_0^{(aa)}$ is small and lies within the limits of ± 50. The quantity $a_0^{(aa)}$ is found from data on rotatory dispersion for denatured proteins. In order to find x_α one has to construct a plot of $\tilde{m}'$ $(\lambda^2/\lambda_0^2 - 1)$ against $(\lambda^2/\lambda_0^2 - 1)^{-1}$. This gives a straight line with slope of $x_\alpha b_0^{(\alpha)}$; the intercept on the ordinate equals $a_0^{(aa)} + x_\alpha a_0^{(aa)}$. Thus, the value of x_α is determined by two independent methods. It should be noted that the presence of aromatic residues has an appreciable effect on the values of the constants in the Moffitt equation. The degree of helicity x_α determined in this way for a number of proteins is in good agreement with the results of other polarimetric determinations (with the aid of one-term or two-term Drude formulas). This quantity agrees also with data obtained by X-ray structure analysis.

The rigorous quantum-mechanical theory of optical activity for homopolymers is concerned with the resonance splitting of the electronic energy levels of monomers into a zone—exciton splitting (see page 155). The calculation of rotational strengths depends on whether the $0 \rightarrow j$ transition under consideration is a "strong" optical transition with large $\boldsymbol{p}_{0j}$ but small $\boldsymbol{m}_{j0}$ or a "weak" transition with small $\boldsymbol{p}_{0j}$ but large $\boldsymbol{m}_{j0}$. For "strong" transitions the exciton contribution can be singled out, which is due to the interaction of excited levels in various monomers. The sum of the exciton rotational

strengths over all monomers is equal to zero—these contributions are "conservative". The non-exciton contributions due to the interaction of level j with other levels, to which transitions are allowed, do not satisfy the condition of conservatism. Figure 5.17 shows typical AORD and CD curves of exciton and non-exciton interactions for conservative and non-conservative contributions. For the α-helix the non-exciton contributions are considerably smaller than the exciton contributions: the theoretical CD curve for an α-helix with

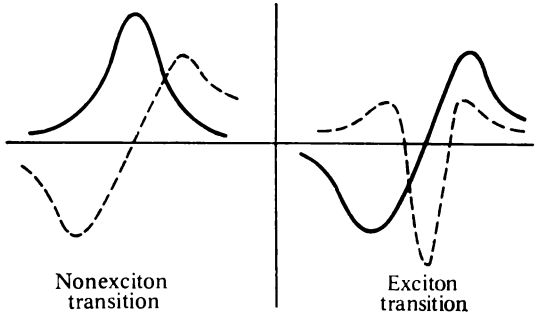


Fig. 5.17. AORD curves (solid lines) and CD curves (dotted lines) for exciton and nonexciton contributions.

a length of 20 turns has three extrema in the region of $\pi \rightarrow \pi^*$ transitions (see page 155) with exciton contributions to R (in 10^{-40} erg·cm^3) equal to —221, 321, —100 (their sum is equal to zero) and non-exciton contributions equal, respectively, to 3.8, —8.6, and —17.1.

Another regular form of the polypeptide chain, the β-form, is in principle also described by the Moffitt formula, but with a small value of b_0. Experiment and calculation give $b_0^{(\beta)} = -30$. A detailed analysis of CD in the region of the $\pi \rightarrow \pi^*$ transition shows that a non-exciton contribution prevails for both the parallel and the antiparallel β-form. Figure 5.18 presents the AORD and CD curves for L-polypeptides in different conformations.

In order to understand the optical activity of nucleic acids, one has to consider the phenomenon of the *induced optical activity* (IOA). Symmetric (i.e., achiral) molecules of dyes, being attached to α-helical polypeptides, show AORD and CD in the regions of absorption. This effect disappears on denaturation of an α-helix-dye complex. The effect is accounted for by the interaction of the dye molecule with a peptide residue near the asymmetric centre. The same behaviour is also shown by the IOA of prosthetic groups and coenzymes. The AORD and CD in the absorption region of *pyridoxal phosphate*,

the coenzyme of *aspartate aminotransferase* (page 200), have furnished information on the structure of the active site of this enzyme. Figure 5.19 presents the AORD curves of deoxyhemoglobin, oxyhemoglobin, and carboxyhemoglobin in the absorption regions of the prosthetic

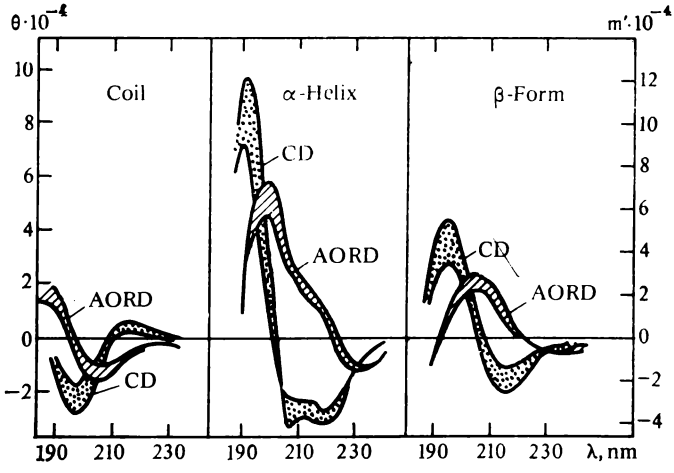


Fig. 5.18. AORD and CD diagrams for L-polypeptides in various conformations

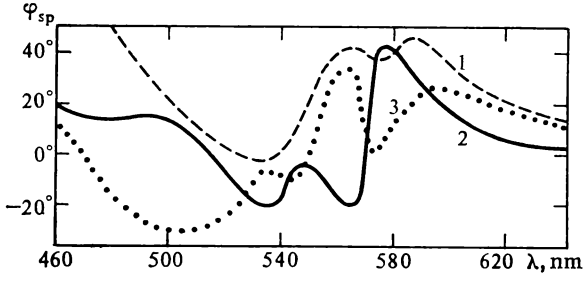


Fig. 5.19. AORD curves for deoxyhemoglobin (1), oxyhemoglobin (2) and carboxyhemoglobin (3).

heme group, which is itself symmetric (see page 53). The chirality of a biopolymer is responsible for the optical asymmetry of the electron shell of the chromophore. In the rigorous theory of induced optical activity one has to deal with the oscillations of atomic nuclei and to arrive at a solution of the electronic-vibrational problem.

The optical activity of nucleosides, nucleotides, DNA, and RNA is characterized by AORD and CD in absorption regions near 260 nm—the absorption bands of nitrogenous bases. These bases are planar and exhibit no chirality, and AORD and CD result from the addition of bases to chiral sugars. Hence, these effects are determined by the induced optical activity. Indeed, the replacement of β-ribose or β-deoxyribose by the corresponding α-compounds changes the sign of the effect.

The rotational strength R_j (5.28) is a tensor rather than a scalar quantity; it has different values in different directions in the molecule. Under ordinary conditions in a liquid one measures the R_j value averaged over all directions. There have been measured the anisotropic values of R_j for complexes of DNA with dyes and antibiotics. The method consists in the orientation of DNA molecules during flow through an optical cell consisting of a set of capillaries. Circular dichroism is measured twice—without orientation when

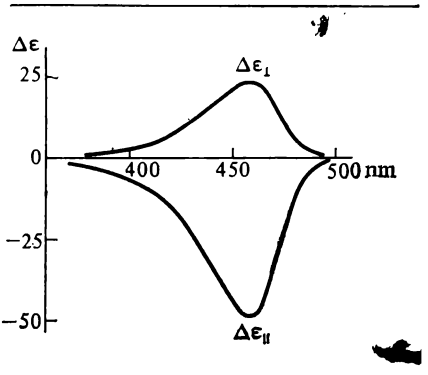


Fig. 5.20. The $\Delta\varepsilon_\perp$ and $\Delta\varepsilon_\parallel$ curves for a DNA-proflavin complex.

$$\Delta\varepsilon = \frac{1}{3}(\Delta\varepsilon_\parallel + 2\Delta\varepsilon_\perp) \tag{5.43}$$

and with orientation of DNA molecules when

$$\Delta\varepsilon_{or} = (1-q)\,\Delta\varepsilon + q\Delta\varepsilon_\parallel \tag{5.44}$$

Here $\Delta\varepsilon$ and $\Delta\varepsilon_{or}$ are the measured amplitudes of CD; $\Delta\varepsilon_\parallel$ and $\Delta\varepsilon_\perp$ are the values of CD along and across the double-helix axis; q is the degree of orientation determined from linear dichroism of DNA. Figure 5.20 shows the $\Delta\varepsilon_\parallel$ and $\Delta\varepsilon_\perp$ curves for a DNA-proflavin complex. It is interesting that these quantities have different signs and, hence, the measured value of $\Delta\varepsilon$ is a small difference of two large values. A study of the anisotropy of the rotational strength yields valuable information on the structure of biopolymers and their complexes.

In an oriented system the symmetry conditions are specific—molecules that have planes of symmetry but are devoid of a centre of symmetry are capable of displaying optical activity. This is *"hidden" optical activity* which reveals only upon orientation. It was discovered in studying DNA and RNA complexes (Makarov and

Poletaev). It has been established that the dye molecules in the complexes are perpendicular to the double-helix axis of nucleic acids.

5.8. Magnetic Optical Activity

In 1845 Faraday wrote in his diary: "... at last I have succeeded in my attempts to magnetize and electrify a beam of light and illuminate a magnetic field line." Thus was discovered the magnetic rotation of the plane of polarization of light propagating along the direction of the magnetic field. This phenomenon was called the *Faraday effect*. What was written by Faraday in his diary has only a figurative meaning—the magnetic field acts not on the light but on the substance that exhibits circular birefringence in the field. The Faraday effect—*magnetic optical rotation* (MOR) and *magnetic circular dichroism* (MCD)—has relatively recently found important applications in molecular biophysics.

The angle of rotation φ defined by formula (5.14) is proportional to the magnetic field strength H (*Becquerel's formula*):

$$\varphi = \frac{\pi}{\lambda}(n_L - n_D)\, l = \mathscr{V} l H \tag{5.45}$$

where $\mathscr{V}$ is the magnetic rotation constant (the *Verdet constant*). In classical electronic theory, the Faraday effect reduces to the *Zeeman effect*. An electron, which is regarded as a harmonic oscillator, oscillates in the absence of a magnetic field with a circular frequency ω_0. In a magnetic field directed along the light beam, a spectral line of frequency ω_0 is split into two, left- and right-circularly polarized components. The splitting is equal to $2 \mid \omega_H \mid$, where ω_H is the circular frequency of the *Larmor precession*:

$$\omega_H = -\frac{eH}{2m_e c} \tag{5.46}$$

(e and m_e are the electron charge and mass and c is the light velocity).

Circular birefringence is defined as follows:

$$\Delta n = n_L - n_D = \left(\frac{\partial n}{\partial \omega}\right)_{H=0} 2 \mid \omega_H \mid = \frac{eH}{m_e c}\left(\frac{\partial n}{\partial \omega}\right)_{H=0} \tag{5.47}$$

or

$$\Delta n = \lambda^2 \left(\frac{\partial n}{\partial \lambda}\right)_{H=0} \frac{eH}{2\pi m_e c^2} \tag{5.48}$$

where λ is the light wavelength. Hence,

$$\varphi = \frac{eH}{2m_e c^2}\, \lambda \left(\frac{\partial n}{\partial \lambda}\right)_{H=0} l \tag{5.49}$$

and

$$\mathscr{V} = \frac{e\lambda}{2m_e c^2}\left(\frac{\partial n}{\partial \lambda}\right)_{H=0} \tag{5.50}$$

We have just obtained Becquerel's formula (5.45). It agrees approximately with experiment in a region far away from absorption bands for diamagnetic substances.

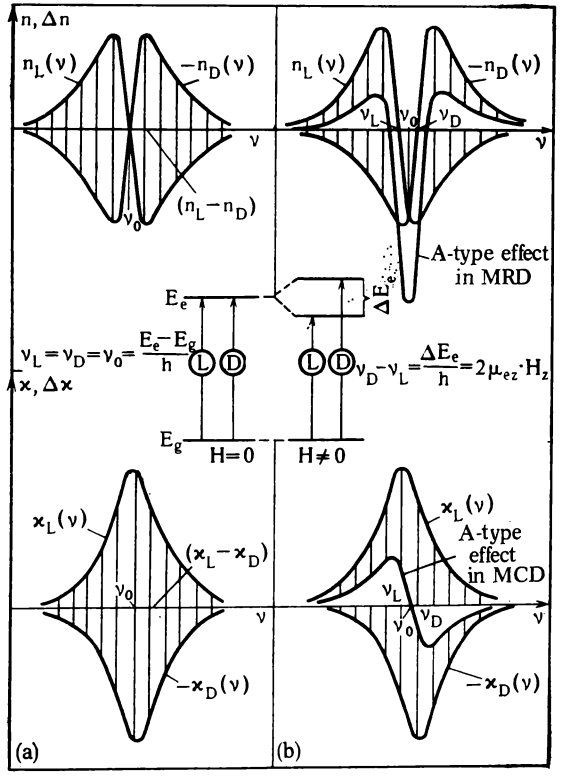
The true theory of the effect can only be constructed on the basis of quantum mechanics. Under the influence of a magnetic field several phenomena occur. First, the ground energy level of the electronic system is split—instead of ω_0 we have $\omega_L = \omega_0 - \gamma_0 H$ and $\omega_D = \omega_0 + \gamma_0 H$, where $\gamma_0 = \omega_H/H$. Second, for the right and left waves there appear different transition probabilities, i.e., extinction coefficients are $\varepsilon_L = \varepsilon\,(1 + bH)$ and $\varepsilon_D = \varepsilon\,(1 - bH)$. Third, the excited levels are also split: $\omega_L = \omega_0 - \gamma_e H$ and $\omega_D = \omega_0 + \gamma_e H$. If the molecules are paramagnetic, then the field H changes the populations of the sublevels of the ground level—the difference in populations gives the factors 1 and $1 - 2\gamma_0 H/\mathrm{k}T$. We obtain ($f\{x\}$ is a function of x)

$$\begin{aligned} \varepsilon_L &= \varepsilon\,(1 + bH)\, f\{\omega - (\gamma_0 + \gamma_e) H\} \\ \varepsilon_D &= \varepsilon\,(1 - bH)\, f\{\omega + (\gamma_0 + \gamma_e) H\}\,(1 - 2\gamma_0 H/\mathrm{k}T) \end{aligned} \tag{5.51}$$

and, expanding $\Delta\varepsilon$ into a series in H and retaining only the linear terms, we find

$$\Delta\varepsilon = 2\varepsilon H\left\{-(\gamma_0 + \gamma_e)\frac{df(\omega)}{d\omega} + \left(b + \frac{\gamma_0}{\mathrm{k}T}\right) f(\omega)\right\} = \\ = A_0 + A_e + B + C/\mathrm{k}T \tag{5.52}$$

The origination of MOR and MCD is illustrated in Fig. 5.21. In the absence of a magnetic field (Fig. 5.21*a*) there are two electronic transitions from the nondegenerate ground level E_0 to a doubly degenerate excited level E_e. The frequencies of the transitions corresponding to the right and left waves and the corresponding refractive indices and extinction coefficients coincide and $n_L(\omega) = n_D(\omega)$ and $\varepsilon_L(\omega) = \varepsilon_D(\omega)$. (The $n_D(\omega)$ and $\varepsilon_D(\omega)$ curves are shown with an opposite sign for convenience.) Figure 5.21*b* explains the appearance of an effect of type A_e, which results from the splitting of the excited level E_e. The $n_L(\omega)$ and $n_D(\omega)$ curves (and, accordingly, the $\varepsilon_L(\omega)$ and $\varepsilon_D(\omega)$ curves) shift symmetrically, as a result of which there appear a symmetrical curve of Δn versus ω in MOR and an asymmetrical curve of $\Delta\varepsilon$ versus ω in MCD. Finally, the effect of type C (Fig. 5.21*c*) arises as a result of the degeneration of the ground state E_0, the elimination of degeneration by the magnetic field and the appearance of the difference in populations of the sublevels in accordance with Boltzmann's law. The C-type effects are asymmetric in MOR and symmetric in MCD.



The rigorous quantum-mechanical theory gives, in the region of absorption, the specific rotation

$$\varphi(0\to j)=-\frac{4\pi N_1}{\hbar c}\left\{\frac{2\omega_j\omega^2[(\omega_j^2-\omega^2)^2-\omega^2\Gamma_j^2]}{\hbar[(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2]^2}A_j+\right.$$
$$\left.+\frac{\omega^2(\omega_j^2-\omega^2)}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2}(B_j+C_j/kT)\right\}H \qquad (5.53)$$

where $\omega_j\equiv\omega_{0j}$; N_1 is the number of molecules in unit volume and (for an isotropic medium)

$$A_j=\frac{1}{6}(m_{jj}-m_{00})\,\mathrm{Im}\,(p_{0j}p_{j0})$$
$$B_j=\frac{1}{3}\,\mathrm{Im}\left\{\sum_{k\neq 0}\frac{m_{k0}}{\hbar\omega_{k0}}p_{0j}p_{jk}+\sum_{k\neq j}\frac{m_{j0}}{\hbar\omega_{kj}}p_{0j}p_{k0}\right\} \qquad (5.54)$$
$$C_j=\frac{1}{6}m_{00}\,\mathrm{Im}\,(p_{0j}p_{j0})$$

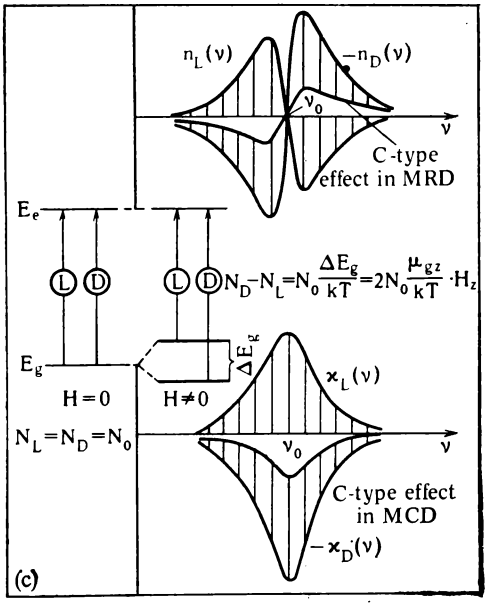


Fig. 5.21. Diagram illustrating the origin of MOR and MCD:
(a) in the absence of a field; (b) the *A*-type effect; (c) the *C*-type effect.

Formula (5.53) transforms to the expression for the contribution of the $0 \to j$ transition to the rotation far away from the region of absorption if we put $\Gamma_j = 0$ in it.

Magnetic circular dichroism is described by the ellipticity

$$\psi(0 \to j) = -\frac{4\pi N_1}{\hbar c}\left\{\frac{4\omega_j\omega^3(\omega_j^2-\omega^2)^2\Gamma_j}{\hbar[\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2]}A_j + \right.$$

$$\left. + \frac{\omega^3\Gamma_j}{(\omega_j^2-\omega^2)^2+\omega^2\Gamma_j^2}(B_j + C_j/kT)\right\} H \qquad (5.55)$$

The Faraday effect is in no way connected with natural optical activity—it is observed in substances composed of both chiral and achiral molecules. When the mirror-reflected light ray is passed again through an optically active substance the rotation is compensated for and disappears, while the Faraday effect is doubled.

Experimental observation of anomalous magnetic rotatory dispersion (AMRD) and magnetic circular dichroism is successfully carried out on systems with a strong magnetic rotation at a relatively weak absorption. Such are, for example, porphyrin compounds—heme-containing proteins, chlorophyll, etc. The effect is proportional to H and therefore it is especially strong in magnetic fields created by superconductors and amounting to 50,000 gauss and upwards. Under ordinary conditions fields of up to 20-30,000 gauss can be produced.

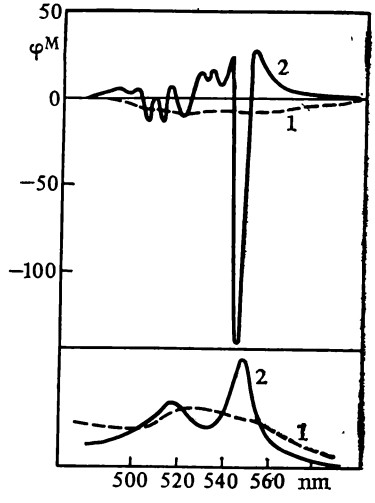


Fig. 5.22. The AMRD curves and absorption spectra of oxidized (1) and reduced (2) forms of cytochrome *c*.

The relative values of MCD for various organic compounds are as follows: about 100 for porphyrins, about 3 for annulenes, about 0.2 for purines, and about 0.00002 for cyclohexanone. The Faraday effect for heme-containing proteins greatly depends on the electronic states of heme and on the conformations of the protein. Figure 5.22 shows the AMRD curves and absorption spectra of native oxidized (ferri-) and reduced (ferro-) cytochrome *c*. The fine structure of MOR of ferricytochrome *c* is clearly seen. No such fine structure is present in the absorption spectrum. The high sensitivity of AMRD and MCD makes it possible to study various cytochromes in suspensions of submitochondrial particles (see page 445). Absorption spectra do not offer such possibilities (Sharonov).

Some of the applications of the Faraday effect to the study of hemoglobin and myoglobin will be considered in Sec. 6.7.

5.9. Vibrational Spectra

The frequencies, intensities, and polarizations of lines and bands in vibrational spectra give information on the structure of molecules. Vibrational spectra are observed either on absorption in the infrared region (IR spectra) or on scattering—Raman spectra. Raman scattering was discovered in 1928 (Raman, Mandelstam, and Landsberg). The possibility of studying molecular vibrations with the aid of spectroscopy in the visible region was very promising, but in ap-

plication to polymers and biopolymers Raman spectra first proved fruitless because of the impossibility of obtaining solutions in an optically pure form, without spurious scattering. For a number of years the vibrational spectra of biopolymers have been studied only as IR absorption spectra; it is only recently that, owing to the development of laser technique, IR spectroscopy has been replaced by Raman spectroscopy.

The theory of molecular vibrational spectra has been worked out in detail (Volkenshtein, Elyashevich, Stepanov, Gribov). It is based on the possibility of studying separately the slow vibrations of atomic nuclei and fast electronic transitions which is proved by the *Born-Oppenheimer theorem*. The problem of the frequencies of nuclear vibrations is classical. It is solved by taking direct account of molecular symmetry. The theory of intensities and polarizabilities in IR and Raman spectra hinges on the *valence-optical scheme*, according to which a particular dipole moment and polarizability can be assigned to each covalent bond. The dipole moment of the molecule is the vector sum of the dipole moments of the bonds, and the polarizability of the molecule is the tensor sum of the bond polarizabilities. The intensity of a particular band in an IR spectrum is determined by a change in the dipole moment of the molecule at a given normal vibration; in the Raman spectrum it is determined by a corresponding change in polarizability.

A molecule consisting of N atoms has $3N - 6$ normal vibrations (a linear molecule has $3N - 5$ normal vibrations). Since the atoms in the molecule are linked to one another, they all participate in each of these vibrations. Situations may occur in which the contribution of a certain covalent bond, atom or group of atoms to a particular normal vibration is greater than that of the remaining bonds and atoms. Such a vibration is *characteristic* of a given atomic group; its frequency and the intensity of the IR band or of the line in the Raman spectrum depend but slightly on the remaining atoms in the molecule. Thereby, the presence of a characteristic spectral band or line is evidence of the presence in the molecule of the corresponding group.

The IR spectra of proteins and polypeptides show the bands of the amide group, i.e., the peptide linkage —CO—NH—. The investigation of low-molecular-weight amides, in particular methyl acetamide (H_3C—CO—NH—CH_3) and related compounds, and theoretical calculations have led to the deduction of the vibrational characteristics of the amide group (Table 5.2).

The shapes of the vibrations listed in Table 5.2 are shown in Fig. 5.23. The IR bands A, B, I, and II are sensitive to hydrogen bonds, whose presence causes a considerable decrease in frequency and expansion of the bands.

In considering the spectra of polymers it is necessary to take into account the interaction of vibrations in individual monomeric

units. An ordered polymer is similar in this sense to a molecular crystal, whose spectrum displays only those vibrations which are in phase in all the elementary cells. The bands of the amide group, in particular, amide I and amide II, prove to be characteristic of the α-helix, β-form, and disordered form of the polypeptide chain. This is shown in Table 5.3.

The spectra of solutions or unoriented films do not give information on vibration polarizations. On the contrary, the polari-

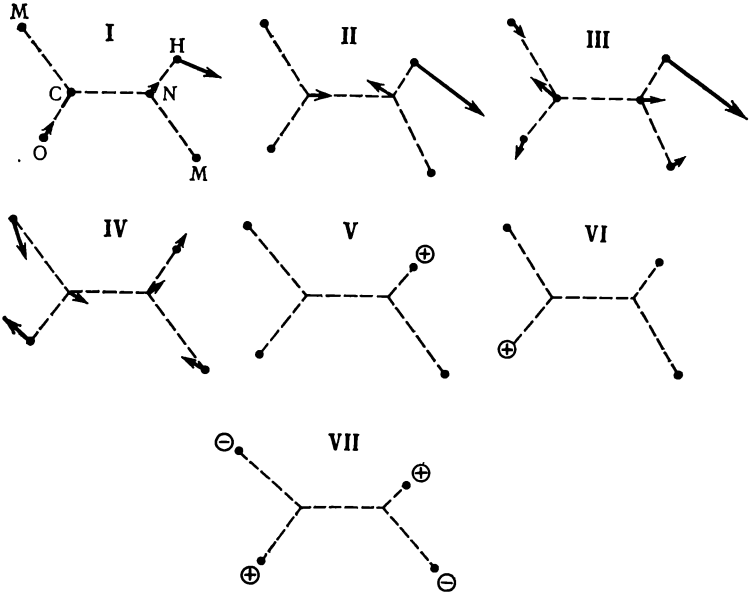


Fig. 5.23. The modes of normal vibrations of the amide group in N-methylacetamide (M is the methyl group).

zation of vibrations is directly observed in investigations of proteins or polypeptides in anisotropic media (in oriented films and fibres) by means of polarized IR radiation. Infrared dichroism is measured, which is characterized by the ratio of the extinction coefficients measured for radiation with the direction of the electric vector of the light wave being parallel and perpendicular to the chosen direction.

Thus, IR spectra can be used for the determination of the secondary structure of a protein, for the study of the ionization of amino-acid residues and, what is especially important, for the study of the kinetics of deuterium exchange.

Table 5.2. The Characteristics of Vibrations of Amide Groups

Band	Frequency, cm^{-1}	Mode of vibration *	Participation of groups, in fractions of unity
Amide A	3300	N—H (ν)	
Amide B	3100		
Amide I	1597-1672	C=O (ν)	C=O (ν) 0.8, C—N (ν) 0.1, —NH (δ) 0.1
Amide II	1480-1575	—NH (δ) in-plane, C—N (ν)	—NH (δ) 0.6, C—N (ν) 0.4
Amide III	1229-1301	C—N (ν), —NH (δ) in-plane	C=O (ν) 0.1 C—N (ν) 0.3, N—H (ν) 0.3
Amide IV	625-767	O=C—N (δ) in-plane	O=C—N (δ) 0.1, O=C—N (δ) 0.4, the rest 0.6
Amide V	640-800	—NH (δ), out-of-plane	
Amide VI	537-606	—CO (δ), out-of-plane	
Amide VII	200	Torsional vibration about the C—N bond	

ν denotes *valence vibration*, i. e., bond stretching.
δ denotes *deformation* or *bond-bending vibration*, in which the valence angle is changed.

Table 5.3. The Band Characteristics of Amide I and Amide II

Conformation	Vibrational mode	Frequency, cm^{-1}	
		amide I	amide II
Disordered	ν_0	1658	1520
Antiparallel	$\nu_\parallel$	1658 (w)	1530 (s)
β-Form	$\nu_\perp$	1632 (s)	1510 (w)
	$\nu_\perp$	1668 (vw)	1550 (w)
Parallel	$\nu_\parallel$	1648 (w)	1530 (s)
β-Form	$\nu_\perp$	1632 (vw)	1530 (w)
α-Helix	$\nu_\parallel$	1650 (s)	1516 (w)
	$\nu_\perp$	1646 (w)	1546 (s)

Notes: (s) means strong, (w) weak, and (vw) very weak bands in the IR spectrum; ∥ and ⊥ are bands corresponding to vibrations of the dipole moment which are parallel and perpendicular to the polymeric chain.

The IR spectroscopy of nucleic acids is less informative and is not practically employed.

The application of IR spectroscopy in the physics of proteins has lately been considerably reduced. Other methods (notably CD) prove to be more efficient for the elucidation of the secondary structure of proteins. Laser Raman spectroscopy allows one to obtain very detailed information on the structure of proteins and other biologically functional substances. There are obtained excellent Raman spectra for solutions, crystalline and amorphous substances, and also for substances in the gaseous phase. The most interesting results are obtained with the aid of *resonance Raman scattering* (RRS). In this case, the wavelength of the scattered laser light enters the region of natural absorption of the chromophore of a biological molecule. The RRS intensities are very high, a fact which allows one to study biological molecules even at very small concentrations. The RRS method has been used to study heme-containing proteins, chlorophyll, enzyme-substrate complexes, etc.

If a molecule does not contain a chromophore required for RRS to be observed, then a chromophore can be introduced as a label by means of chemical or nonchemical binding.

5.10. Nuclear Magnetic Resonance and Electron Paramagnetic Resonance

The spectra of nuclear magnetic resonance and electron paramagnetic (or spin) resonance (NMR and EPR or ESR) are widely employed in biophysics. Let us consider the essence of magnetic resonance.

We place a particle with a magnetic moment μ in a uniform magnetic field of strength H_0. The magnetic moment will experience precession about the field direction (Fig. 5.24*a*) with a frequency proportional to H_0 (cf. formula 5.46):

$$\omega_0 = \gamma H_0 = \frac{\mu}{p} H_0 \tag{5.56}$$

where p is the angular momentum of the particle (say, the electron spin). We subject the precessing particle to the action of a low variable magnetic field H_1 directed at right angles to H_0. The linearly polarized field H_1 can be split into two circular components (page 162). One of these components coincides in direction with the precession. If the vibrational frequency of H_1 coincides with ω_0, a resonance appears, a strong absorption of the H_1 wave. This phenomenon was discovered by Zavoisky in 1945.

Such is the elementary classical description of the effect. Quantum mechanics shows that the energy levels of a particle possessing a magnetic moment are split in a magnetic field. We shall begin with

NMR. Atomic nuclei whose mass or atomic number is odd have a nuclear spin and, hence, a nuclear magnetic moment. The nuclear spin is equal to zero for ^{12}C, ^{16}O, etc., and is different from zero for H, D, ^{13}C, ^{19}F, etc.

If I is the nuclear spin number, the corresponding magnetic moment is equal to $[I(I+1)]^{1/2}\gamma\hbar$, where $\gamma\hbar = g\mu_0$ is the *gyromagnetic ratio*, g is the *splitting factor*, and μ_0 is the *nuclear magneton*. In the magnetic field H_0, as a result of a Zeeman splitting (see page 174), there appear $2I + 1$ levels with energies

$$E = -\gamma\hbar H_0 m$$
$$(m = I, I-1, \ldots, -I) \quad (5.57)$$

The distance between neighbouring levels equals $\gamma H_0\hbar$. For a proton $I = 1/2$ and in the field H_0 two levels appear, which correspond to the parallel and antiparallel directions of the spin (Fig. 5.24*b*). Resonance occurs at

$$\hbar\omega_0 = \gamma\hbar H_0 = g\mu_0 H_0 \quad (5.58)$$

where

$$\mu_0 = \frac{e\hbar}{2m_p c} = 5.05 \times 10^{-24} \text{ erg/G} \quad (5.59)$$

where e is the charge of the proton and m_p is its mass. The g-factor for the proton is equal to 5.58. If $H_0 = 10{,}000$ gauss, the resonance frequency for protons is 42.6 MHz, which corresponds to the radio-wave range.

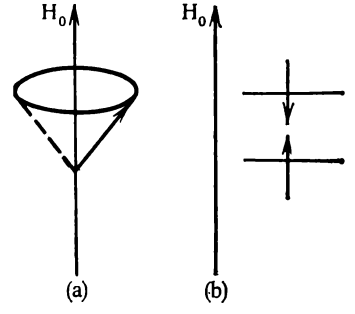


Fig. 5.24. Schematic representation of paramagnetic resonance:
(*a*) the classical model and (*b*) the quantum-mechanical model.

At a given temperature T the number of protons at the lowest level is larger than that at the upper level. The ratio of the populations of the upper and lower levels at equilibrium is expressed by the Boltzmann factor (see page 175):

$$\exp(-\hbar\omega_0/kT) = \exp(-g\mu_0 H_0/kT) \approx 1 - g\mu_0 H_0/kT \quad (5.60)$$

The exponential factor is replaced by the first two expansion terms, since $g\mu_0 H_0 \ll kT$ at $T = 300$ K and ordinary values of H_0. At $H_0 = 10^4$ gauss the Boltzmann factor is $1\text{-}14 \times 10^{-6}$.

When a radiofrequency field H_1 is switched on, there occur transitions from the lower to the upper level (absorption) and vice versa (emission). If the probabilities of both processes are equal, the levels must be rapidly saturated—their populations become the same and absorption is stopped. This is not actually observed, however, since

the nuclear spins are capable of giving off their energy even without radiation. A relaxation process takes place, which makes the spin system revert to the equilibrium state, to which there corresponds the Boltzmann distribution. It results from the interaction of nuclear spins with the *lattice*, i.e., with other nuclei in a state of thermal motion. When the field H_1 is turned off, the energy released is converted into thermal energy of the lattice. The change in level populations that occurs after the field H_1 has been switched off is described by the equation

$$\Delta n\,(t) = \Delta n\,(0)\,\exp\,(-t/T_1) \tag{5.61}$$

where $\Delta n\,(t)$ is the excess of protons at time t at the upper level as compared with the equilibrium population; $\Delta n\,(0)$ is the excess of protons at the moment when the field is switched off ($t = 0$); T_1 is the *spin-lattice relaxation time*. The time T_1 depends on the concentration of magnetic nuclei in the substance, on the mobility of molecules and temperature. In crystals T_1 is of the order of minutes; in gases and liquids it is of the order of seconds and less. In the presence of paramagnetic impurities T_1 can be reduced to 10^{-4} s. The width of the NMR spectral line is the larger, the shorter is the lifetime of the nucleus at a given energy level. The spin-lattice relaxation makes a contribution to the linewidth of the order of T_1^{-1}.

There is a second process—the *spin-spin interaction* of magnetic nuclei. Apart from the field H_0, each of the spins is acted on by a local field created by the neighbouring nuclei and equal to

$$H_{\text{loc}} = \frac{\mu}{r^3}\,(3\cos^2\theta - 1) \tag{5.62}$$

where r is the separation of the nucleus from a given point; θ is the angle between μ and r. The resonance condition therefore has the form

$$\omega_0 = \gamma\,(H_0 + H_{\text{loc}}) \tag{5.63}$$

rather than the form of (5.56), which is valid for an isolated spin. The order of the quantity H_{loc} is a few gauss. The spin-spin interaction makes also a contribution to the linewidth due to the field inhomogeneity and the energy exchange between the protons, which is determined by this interaction. The *spin-spin relaxation time* T_2 in solids is much longer than T_1 and the linewidth is determined only by T_2. In nonviscous liquids T_2 is of the same order of magnitude as T_1 and the linewidth is fractions of Hz.

The NMR spectra for liquids and solutions contain a series of narrow lines corresponding to structurally nonequivalent protons. For example, the spectrum of 1,1,2-trichloroethane, $ClH_2C{-}CHCl_2$, at moderate resolution shows two lines corresponding to the protons of the CH_2 and CH groups with an intensity ratio of 2 : 1. This occurs due to the screening effect of the electron environment of the

nucleus on its spin. The electrons precess in a direction opposite to the nuclear precession and create a secondary field $H' = -\sigma H_0$. The nucleus is acted on by an effective field:

$$H_{\text{eff}} = H_0 + H' = (1 - \sigma) H_0 \qquad (5.64)$$

where σ is the *screening constant*. Its value is different in different directions in the molecule, i.e., σ is an anisotropic quantity. In a liquid it is averaged.

The position of the NMR spectral line relative to a certain reference line is called the *chemical shift*. The usual standard compound for the proton resonance of organic compounds is tetramethylsilane, $(CH_3)_4Si$, and for aqueous solutions of biopolymers 2,2-dimethyl-2-silanepentane-5-sulphonic acid is used as the standard. The chemical shift is expressed by the ratio of the shift of the field frequency to the standard value multiplied by 10^6:

$$\delta = \frac{\Delta\omega}{\omega_0} \cdot 10^6 = \frac{\Delta H}{H_0} \cdot 10^6 \qquad (5.65)$$

In biopolymers, the value of δ ranges from -0.5 to -2.0 for aliphatic protons and from -6.0 to -8.5 for aromatic protons, etc.

At high resolution the NMR lines show a *hyperfine* (multiplet) *structure*. It arises from an internuclear magnetic interaction transmitted through bonding electrons, i.e., an indirect spin-spin interaction. For instance, in $ClH_2C—CHCl_2$ the proton of the CH group may be in two states—one with spin $+\frac{1}{2}$ and the other with spin $-\frac{1}{2}$. Therefore, the line of the protons of the neighbouring CH_2 group is split into two lines. In the CH_2 group there are possible three nonequivalent states of a pair of protons: $+\frac{1}{2}, +\frac{1}{2}$; $+\frac{1}{2}, -\frac{1}{2}$; $-\frac{1}{2}, +\frac{1}{2}$; $-\frac{1}{2}, -\frac{1}{2}$. The line of the proton of the CH group experiences a triplet splitting (Fig. 5.25).

The study of chemical shifts and the hyperfine structure yields quantitative information on the arrangement and interactions of atomic nuclei. Using high-resolution instruments (powerful superconducting magnets, the frequencies being of the order of 300 MHz upwards) one can resolve a multitude of overlapping proton resonance lines in proteins. Here the selective deuteration of the protein is also essential. The methods of NMR have been successfully employed for investigations of the conformational properties of biopolymers, interactions of proteins with pharmacological substances, enzyme-ligand, antigen-antibody, and other types of interaction. The possibilities of NMR spectroscopy in molecular biophysics are unlimited.

Nuclear magnetic resonance on ^{31}P and ^{13}C is also effectively used for the study of living cells and tissues. It is possible to separate

weak signals from the background with the aid of the methods of Fourier transformation. For this purpose, use is made of a powerful radiofrequency pulse which excites simultaneously the entire resonance spectrum, whose individual frequencies are sorted out by means of computers. This method has been used to study phenomena in a living muscle—the changes in the concentration of ATP and creatine

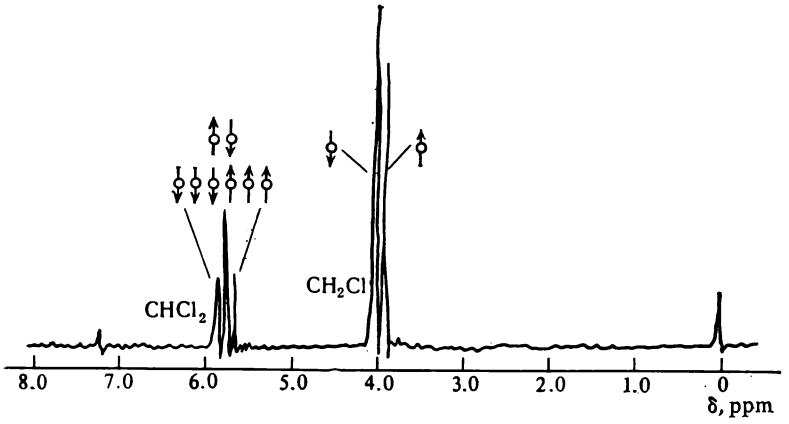


Fig. 5.25. The hyperfine structure of the proton resonance spectrum of $ClH_2C—CHCl_2$.

phosphate during muscle contraction (see Sec. 12.3). The position of the ^{31}P peak of inorganic phosphate in the cardiac muscle depends on pH and is therefore shifted in the case of cardiac insufficiency. This opens up wide possibilities of using NMR in cardiology.

We shall now consider electron paramagnetic resonance (EPR), which is also called electron spin resonance (ESR). An electron has spin $\frac{1}{2}$; its energy level in a field H_0 is split into two levels with a distance between them corresponding to the resonance condition (cf. formula 5.58)

$$\hbar\omega_0 = g\mu_B H_0 \tag{5.66}$$

where μ_B is the Bohr magneton which exceeds the nuclear magneton by 1836 times:

$$\mu_B = \frac{e\hbar}{2m_e c} = 0.93 \times 10^{-20} \text{ egr/G} \tag{5.67}$$

where m_e is the electron mass. For a free electron $g = 2.0023$. In the field $H_0 = 10^4$ gauss the resonance frequency $\nu_e = 2.8 \times 10^{10}$ s^{-1} ($\nu_p = 4.26 \times 10^7$ s^{-1}).

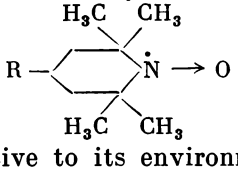
In the overwhelming majority of organic compounds the electron

spins are compensated and no ESR is present. ESR is observed in *free radicals* and in molecules with an odd number of electrons. Here ESR gives essential information on the electronic structure of the system. The ESR lines undergo hyperfine splitting due to a local field created by the magnetic moments of nuclei. For example, ^{14}N has a magnetic moment of unity and, hence, the projections of the nuclear spin on the direction of the field H_0 correspond to the values of the magnetic quantum number $m = 1, 0, -1$. The local field that acts on the electron of a free radical that is near the nucleus of ^{14}N has three values and the ESR peak is split into a triplet.

A so-called *fine structure* appears in crystals due to the *g-factor anisotropy*. Instead of a single line there are observed a group of lines, whose positions and intensities depend on the orientation of the field H_0 relative to the crystal axes. In liquids no fine structure is observed.

The linewidth in the ESR spectrum is determined, just as in NMR, by the spin-spin and spin-lattice relaxation. The times T_1 and T_2 depend on the mobility of particles with nonzero spins (T_2) and on the mobility of surrounding molecules (T_1). The spin-spin relaxation time T_2 is nearly independent of temperature and is determined by the concentration of paramagnetic particles. The time T_1 increases as the temperature drops.

Electron spin resonance is employed for the solution of many problems of biology and biophysics. Free radicals are formed in a number of enzymatic redox processes; under the action of radiation, etc. ESR provides very valuable information on these processes. ESR is very important for the study of proteins that contain, as cofactors, metal atoms with uncompensated spin (hemoproteins; see Sec. 6.8, etc.). The ESR studies of biopolymers based on the *spin-label techniques* have found wide application. To a biopolymer there is added a stable free radical containing an unpaired electron. Especially convenient are iminoxyl radicals



The spin-label is sensitive to its environment; its ESR spectrum depends on the conformational state of a biopolymer. The mobility of the label is associated with the local mobility of the biopolymer and with the mobility of its molecule as a whole. Important specific features of the structure and dynamics of protein molecules have been established through the use of spin-labels. The probing of a molecular or supermolecular biological system with the aid of spin-levels is similar to the use of luminescence labels (page 161). Often the two methods are profitably combined.

Chapter 6 Physics of Enzymes

6.1. Chemical Kinetics and Catalysis

The most important function of proteins is to act as enzymes. Enzymes serve as catalysts in all biological reactions. A *catalyst* accelerates the rate of a reaction but is not consumed in the overall process.

The rates of chemical reactions usually greatly depend on temperature. Empirically, this dependence is expressed by the Arrhenius formula for the reaction rate constant (see page 22):

$$k = A \exp(-E^{\ddagger}/RT) \tag{6.1}$$

where $E^{\ddagger}$, the *energy of activation*, characterizes the energy barrier that must be overcome by the system for the reaction to occur. The dependence of k on $E^{\ddagger}/RT$ follows from the Boltzmann distribution among energy levels; the exponential indicates the fraction of molecules that possess sufficient energy $E^{\ddagger}$ for the reaction to take place. The constants k determine the reaction rates. For a reversible reaction

$$\mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{B}$$

the rates are given by ([A] and [B] are the concentrations):

$$\vec{v} = k_1 [\mathrm{A}], \quad \text{and} \quad \overleftarrow{v} = k_{-1} [\mathrm{B}] \tag{6.2}$$

At equilibrium $\vec{v} = \overleftarrow{v}$ and the equilibrium constant is defined by

$$K = \frac{[\mathrm{B}]_{\mathrm{eq}}}{[\mathrm{A}]_{\mathrm{eq}}} = \frac{k_1}{k_{-1}} \tag{6.3}$$

The subscript "eq" indicates an equilibrium concentration. The free-energy difference between reaction products and reactants is equal to

$$\Delta G = -RT \ln K = \Delta H - T\Delta S \tag{6.4}$$

where ΔH is the enthalpy change and ΔS is the entropy change. The reaction is possible only if $\Delta G < 0$, i.e., if the free energy is decreased. However, this condition is necessary but is still insufficient

for the reaction to occur. The energy of activation $E^{\ddagger}$ may be so great that the rate constant is vanishingly small and the reaction does not proceed. The catalyst serves to lower the activation barrier.

The meaning of the Arrhenius formula is disclosed in the *theory of absolute reaction rates* (the transition-state or activated-complex theory) proposed by Eyring.

It is assumed that the occurrence of the reaction does not upset the Boltzmann distribution of molecules over states with different energies. For the reaction to proceed it is necessary that an energy barrier be surmounted. Figure 6.1 shows the profile of free energy for a certain process: $\Delta G = G_1 - G_2$ is the free-energy difference between the initial state 1 and the final state 2; $G^{\ddagger}$ is the free energy of activation in the process $1 \rightarrow 2$. Let us find the number of systems which are in a certain interval δ of the *reaction coordinate* ξ at the top of the energy barrier. The reaction rate is the number of systems that pass over the barrier per unit time. The mean velocity of the system is given by

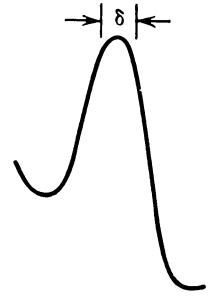


Fig. 6.1. The free-energy profile for a chemical reaction.

$$\bar{v} = \frac{\int_0^\infty v \exp(-mv^2/2kT)\, dv}{\int_0^\infty \exp(-mv^2/2kT)\, dz} = \left(\frac{kT}{2\pi m}\right)^{1/2} \tag{6.5}$$

We have made use of the Maxwell-Boltzmann distribution over kinetic energies, $mv^2/2$. The reaction rate, i.e., the number of jumps over the energy barrier per unit time is given by

$$\nu = \frac{c'}{\delta}\bar{v} = \frac{c'}{\delta}\left(\frac{kT}{2\pi m}\right)^{1/2} \tag{6.6}$$

where c' is the number of systems in unit volume within the interval δ. At the same time we have

$$\nu = kc_A c_B \ldots \tag{6.7}$$

where c_A, c_B, ... are the concentrations of the reactants. From Eqs. (6.6) and (6.7) it follows that

$$k = \frac{c'}{c_A c_B \ldots}\left(\frac{kT}{2\pi m}\right)^{1/2}\frac{1}{\delta} = K'\left(\frac{kT}{2\pi m}\right)^{1/2}\frac{1}{\delta} \tag{6.8}$$

Here K' is the equilibrium constant for the passage of reactants A, B, . . . to an activated state. The equilibrium constant is expressed n terms of partition function (sums over states):

$$Z = \sum_i \exp(-E_i/\mathrm{k}T) \tag{6.9}$$

where E_i is the energy of the ith state of the system. Summation is carried out over all states. In our case,

$$K' = \frac{Z_{ac}}{Z_A Z_B \ldots} \tag{6.10}$$

where Z_{ac}, Z_A, Z_B, . . . are the partition functions for the activated complex (the transition state) and the reactants A, B, . . ., respectively. For the translational motion of a particle the set of states is continuous; the partition function (6.9) is replaced by an integral, which is equal to

$$Z_{tr} = \delta^3 \left(\frac{2\pi m\mathrm{k}T}{h^2}\right)^{3/2} \tag{6.11}$$

where h is Planck's constant and m is the mass of the particle. We assume the volume occupied by the particle to be equal to δ^3; Z_{tr} is fhe partition function with respect to three translational degrees of treedom. But as the system has only one degree of freedom—along the direction of the reaction, we obtain:

$$Z_{ac} = Z'_{ac}\delta \left(\frac{2\pi m\mathrm{k}T)}{h^2}\right)^{1/2} \tag{6.12}$$

where Z'_{ac} is the partition function for the activated complex over all degrees of freedom, except the translational degree of freedom. From Eqs. (6.8), (6.10), and (6.12) we obtain:

$$k = \frac{\mathrm{k}T}{h} \frac{Z'_{ac}}{Z_A Z_B \ldots} \tag{6.13}$$

The conditional length δ is reduced. The statistical quantities are directly associated with thermodynamic quantities. We have

$$Z = \exp(-G/RT) \tag{6.14}$$

and, hence,

$$k = \frac{\mathrm{k}T}{h} \frac{\exp(-G'/RT)}{\exp(-G_A/RT)\exp(-G_B/RT)\ldots} =$$

$$= \frac{\mathrm{k}T}{h} \exp\left(-\frac{G' - G_A - G_B - \ldots}{RT}\right) = \frac{\mathrm{k}T}{h} \exp\left(-\frac{G^{\ddagger}}{RT}\right) \tag{6.15}$$

Here $G^{\ddagger}$ is the excess free energy of the activated complex as compared with the total free energy of the reactants, i.e., the *free energy of activation.*

In the derivation of expression (6.15) it is assumed that the system, having reached the activation barrier, is necessarily made to undergo a chemical transformation. This may not take place, however. Introducing an additional factor—the transmission coefficient $\varkappa \leqslant 1$, we obtain the final expression for the rate constant of a homogeneous gaseous reaction—the *Eyring formula*:

$$k = \varkappa \frac{kT}{h} \exp(S^{\ddagger}/R) \exp(-H^{\ddagger}/RT) \tag{6.16}$$

where $S^{\ddagger}$ is the entropy of activation and $H^{\ddagger}$ is the enthalpy of activation. The value of $\varkappa = 1$ corresponds to an adiabatic reaction, i.e., to a process in which the parameters of the mechanical system change slowly. In a chemical reaction such parameters are the positions of atomic nuclei. With each of the configurations of the nuclei the electrons move as though the nuclei remain fixed. If the process is nonadiabatic, $\varkappa \ll 1$ and may assume values of the order of 10^{-5}. The $\varkappa$ values observed in gas-phase reactions are mostly close to unity, i.e., these reactions proceed adiabatically. The factor kT/h at ordinary temperatures is of the order of 10^{13} s^{-1}.

Let us compare the Eyring formula (6.16) with the Arrhenius formula (6.1). If the enthalpy of activation is identified with the activation entropy, the pre-exponential term will be given by

$$A = \varkappa \frac{kT}{h} \exp(S^{\ddagger}/R) \tag{6.17}$$

The values of the activation parameters are determined from the temperature dependence of k. At $\varkappa = 1$ we have

$$\ln k = \frac{S^{\ddagger}}{R} + \ln \frac{kT}{h} - \frac{H^{\ddagger}}{RT} \tag{6.18}$$

The second term depends but slightly on T and may therefore be neglected. We obtain the linear dependence of $\ln k$ on $1/T$. From the slope of the straight line we find $H^{\ddagger}$; the value of $S^{\ddagger}$ is found from the ordinate intercept. The nonlinearity of the dependence of $\ln k$ on T^{-1} is an indication of the complexity of the process, i.e., it shows that the process becomes cooperative. In such a process the free energy of activation is not constant and depends on the number of systems that have reacted and, hence, on T. Suppose we are riding in an overcrowded bus and it is time for us to get off. At first it is difficult to do so: a large energy of activation is required. As the number of people that have left the bus increases, this energy is diminished. This is a spectacular example of cooperative kinetics.

Calculations of rate constant by formula (6.16) can be carried out only for the simplest gas-phase reactions.

It should be emphasized that h/kT must not be treated as the lifetime of the activated complex because it leads to a contradiction:

a statistical equilibrium cannot be set up in 10^{-13}-10^{-14} s. The lifetime of the system in a state corresponding to the values of reaction coordinate from ξ to $\xi + \delta$ is actually proportional to δ. This time must be treated in the same way as the lifetime of states dealt with in the Maxwell-Boltzmann theory. From the smallness of h/kT one cannot conclude that the equilibrium between the activated complex and the reactants cannot be established. The term "activated complex" refers to the states of an atomic system that undergoes transformations and does not imply the existence of a metastable complex, which can be studied by chemical and physical methods.

The Eyring theory is applicable to gas-phase chemical reactions and also to physical kinetic processes (viscous flow, diffusion, etc; cf. page 427). This theory cannot, however, be used in dealing with reactions taking place in solutions.

Let us now turn to catalysis. As has already been said, the role of a catalyst is to lower the free energy of activation and, hence, to increase the rate constant of the reaction. Two types of catalysis are distinguished: *heterogeneous* and *homogeneous*. In the first case, the catalyst forms a separate phase, and the reaction takes place at the interface between the phases. The process begins with *adsorption* of the reactants at this interface, say on the surface of a solid catalyst, which speeds up the reaction in a gas or a liquid. The adsorption is accompanied by a change in the electronic structure of the reactants and by a decrease in the free energy of activation. After the reaction has gone to completion the products are desorbed from the surface. In homogeneous catalysis the catalyst and reactants are in a single phase, say in a solution. The catalyst forms an intermediate compound with the reactants which is then decomposed into the catalyst and reaction products.

Enzymes function either in solution or in supermolecular structures. The sorption of the reactants, called *substrates*, and the reaction take place on some part of the surface of a large protein molecule. In this sense, enzyme catalysis is similar to heterogeneous catalysis. But the protein-enzyme and the small molecules of the substrates are present in one phase in solution. There is a strict stoichiometry of the interaction—as a rule, one protein globule interacts with one molecule of the substrate or another ligand. The interaction leads to the formation of an *enzyme-substrate complex* (ESC), whose structure and properties are studied by physical and chemical methods. Enzyme catalysis in solution is a homogeneous catalysis.

We may regard an enzyme as a "black box" that transforms the input signal, the substrate molecule, to the output signal, the *product* molecule. There are two methods of investigating the arrangement and operation mechanism of the "black box": the study of the molecular structure of the enzyme and enzyme-substrate complex and the study of the kinetics of enzymic reactions.

6.2. Kinetics of Simple Enzymatic Reactions

Let us consider the simplest enzymic reaction involving the conversion of the substrate to the product in a one-step process:

$$F_0 + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} F_1 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} F_0 + P$$

Here F_0 is the free enzyme; S is the substrate; F_1 is the enzyme-substrate complex (ESC); P is the reaction product. The reaction rate constants are indicated above and below the arrows. Assume, for the sake of simplicity, that we may ignore k_{-2}, i.e., the reverse reaction. The kinetic equations have the form (here F_i, S, and P are the concentrations):

$$\begin{aligned} \dot{S} &= -k_1 F_0 S + k_{-1} F_1 \\ \dot{F}_1 &= k_1 F_0 S - (k_{-1} + k_2) F_1 \\ \dot{P} &= -\dot{S} = k_2 F_1 \end{aligned} \tag{6.19}$$

The total concentration of enzyme is constant:

$$F_0 + F_1 = E = \text{constant} \tag{6.20}$$

Eliminating F_0 from the second equation of (6.19) with the aid of Eq. (6.20), we obtain:

$$\dot{F}_1 = k_1 ES - (k_{-1} + k_2 + k_1 S) F_1 \tag{6.21}$$

The system is in a steady state if there is a large excess of the substrate, $S \gg E$. As long as the stationary state is retained $\dot{F}_1 = 0$ (and, of course, $\dot{F}_0 = 0$). We shall illustrate this by an example. Suppose we are to lace up our shoes. To do this quickly, we place our foot on a chair, which plays the role of a catalyst. The substrate is the unlaced-up shoe and the product is the shoe that has been laced up; the enzyme is the free chair and the enzyme-substrate complex is the foot placed on the chair. If a large number of people are to lace up their boots (say, sailors called upon to do so by the alarm), and the number of chairs is small, then the number of chairs on which the feet are placed will be constant for a certain period of time.

Thus, in the steady state $\dot{F}_1 = 0$ and from Eq. (6.21) we get:

$$F_1 = \frac{k_1 ES}{k_{-1} + k_2 + k_1 S} \tag{6.22}$$

From Eq. (6.19) it follows that

$$v = \dot{P} = -\dot{S} = k_2 F_1 = \frac{k_1 k_2 SE}{k_{-1} + k_2 + k_1 S} \tag{6.23}$$

where v is the reaction rate (or the reaction velocity). Dividing the numerator and the denominator by k_1, we arrive at the *Michaelis-Menten equation*:

$$v = \frac{k_2 S E}{K_M + S} \tag{6.24}$$

where the *Michaelis constant* is defined as

$$K_M = \frac{k_{-1} + k_2}{k_1} \tag{6.25}$$

Before we consider this equation let us prove that the stationary state is actually realized at $S \gg E$. We integrate Eq. (6.21), assuming S to be practically constant, i.e., S is treated as a slowly changing parameter. At time $t = 0$ when the enzyme is added to the substrate $F_1(0) = 0$ (the enzyme present is free). We have

$$F_1 = \frac{k_1 E S}{k_{-1} + k_2 + k_1 S} (1 - e^{-t/\tau}) \tag{6.26}$$

where

$$\tau = (k_{-1} + k_2 + k_1 S)^{-1}$$

The enzyme-substrate complex F_1 relaxes to a steady-state value corresponding to $\dot{F}_1 = 0$ with the time constant τ the more rapidly, the larger is the concentration of the substrate S. From Eq. (6.26) it follows that

$$\dot{F}_1 = k_1 E S \exp(-t/\tau) \to 0 \tag{6.27}$$

Substituting the value of F_1 from Eq. (6.26) and that of $\dot{F}_1$ from Eq. (6.27) into the first equation of (6.19), we find that

$$\dot{S} = -k_1 E S \exp\left(-\frac{t}{\tau}\right) - \frac{k_1 k_2 E S}{k_{-1} + k_2 + k_1 S}\left[1 - \exp\left(-\frac{t}{\tau}\right)\right] \tag{6.28}$$

that is, at $t \to \infty$ the value of $\dot{S}$ actually tends to a steady-state value (6.23). This is associated with the initial assumption that S changes slowly. The relative rate of change of S is really slow at $S \gg E$. The quantity $\dot{S}$ decreases monotonously with time, its maximum value corresponding to the initial moment $t = 0$:

$$\dot{S}(0) = -k_1 E S$$

The maximum measure of the change of S is therefore

$$\Delta S \approx \dot{S}(0)\,\tau = -\frac{k_1 E S}{k_{-1} + k_2 + k_1 S} \tag{6.29}$$

It is easy to see that at $E \ll S$

$$|\Delta S| \ll S$$

At the same time $\Delta F_1 \approx \dot{F}_1(0)\,\tau \approx -\Delta S$ and $\Delta F_1 / E \gg |\Delta S| / S$.

Let us now return to the Michaelis-Menten equation (6.24). The dependence of v on S is similar to the Langmuir isotherm (Fig. 6.2). The curve has no singularities; the reaction rate tends asymptotically to its maximum value at $S \to \infty$:

$$v_{max} = k_2 E \tag{6.30}$$

Hence,

$$v = \frac{v_{max} S}{K_M + S} \tag{6.31}$$

At $v = 0.5 v_{max}$ we have $S = K_M$, i.e., the Michaelis constant is the substrate concentration at which the reaction rate is equal to half the maximum value.

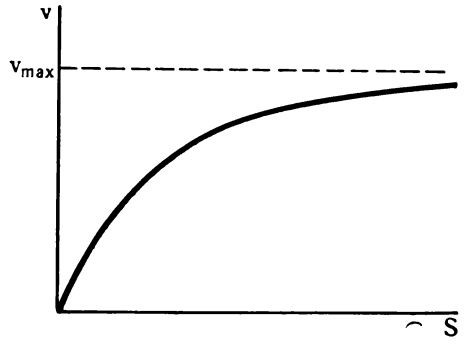


Fig. 6.2. The Michaelis-Menten curve.

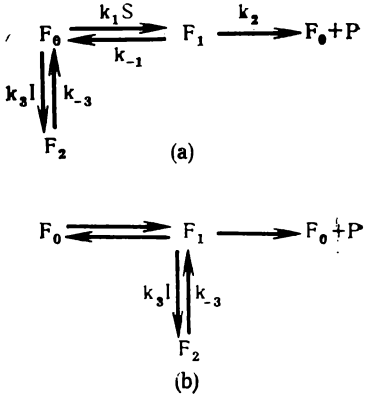


Fig. 6.3. Schematic representation of processes with competitive (*a*) and noncompetitive (*b*) inhibition.

Equation (6.31) can be conveniently transformed to the Lineweaver-Burk equation

$$\frac{1}{v} = \frac{K_M}{v_{max}} \frac{1}{S} + \frac{1}{v_{max}} \tag{6.32}$$

Plotting a graph of $1/v$ against $1/S$, we find $1/v_{max}$ from the point of intersection of the straight line with the ordinate and then the constant K_M from the slope of the straight line. The Michaelis constant can be represented in the form:

$$K_M = k_2/k_1 + K_S$$

where

$$K_S = k_{-1}/k_1$$

All the three rate constants can be determined if Eq. (6.32) is transformed with the aid of Eq. (6.30) to the form

$$\frac{v_{max}}{v} - 1 = \frac{K_S}{S} + \frac{v_{max}}{k_1 ES} \tag{6.33}$$

At a given value of S the dependence of $v_{max}/v - 1$ on v_{max} is represented by a straight line that makes an intercept K_S/S on the ordinate and an intercept $-1/k_1 ES$ on the abscissa. From the straight lines obtained at different values of S and E we can find k_1, k_{-1} and k_2.

We shall now consider the same reaction in the presence of a modifier, a substance which brings about a change in reaction velocity. If the modifier slows down the reaction, it is called an *inhibitor*; otherwise, it is known as an *activator*. We designate the inhibitor concentration by I. Figure 6.3 presents graphs for the two simplest cases of inhibition. In the case of *competitive inhibition* an enzyme forms, apart from an enzyme-substrate complex, an inactive complex with I denoted by F_2. The inhibitor competes with the substrate, being bound with the active site of the enzyme. In *noncompetitive inhibition* the complex F_2 is formed upon interaction of the enzyme-substrate complex F_1 with I.

The equations of steady-state kinetics at $S \gg E$ and $I \gg E$ have the following form in the first case:

$$\begin{aligned}
&\dot{F}_0 = -(k_1 S + k_3 I) F_0 + (k_{-1} + k_2) F_1 + k_{-3} F_2 = 0 \\
&\dot{F}_1 = k_1 S F_0 - (k_{-1} + k_2) F_1 = 0 \\
&\dot{F}_2 = k_3 I F_0 - k_{-3} F_2 = 0 \\
&E = F_0 + F_1 + F_2, \quad v = k_2 F_1
\end{aligned} \tag{6.34}$$

Their solutions have the form

$$F_1 = \frac{k_1 S}{k_{-1} + k_2} F_0, \quad F_2 = \frac{k_3 I}{k_{-3}} F_0, \quad E = \left(1 + \frac{k_1 S}{k_{-1} + k_2} + \frac{k_3 I}{k_{-3}}\right) F_0$$

and, hence,

$$v = \frac{k_2 k_1 S}{k_{-1} + k_2} F_0 = \frac{k_2 S E}{K_M + S + K_M K_I I} \tag{6.35}$$

where $K_I = k_3/k_{-3}$. The maximum reaction velocity $v_{max} = k_2 E$ does not change, but the velocity at small S falls off.

For noncompetitive inhibition

$$\begin{aligned}
&\dot{F}_0 = -k_1 S F_0 + (k_{-1} + k_2) F_1 = 0 \\
&\dot{F}_1 = k_1 S F_0 - (k_{-1} + k_2 + k_3 I) F_1 + k_{-3} F_2 = 0 \\
&\dot{F}_2 = k_3 I F_1 - k_{-3} F_2 = 0 \\
&E = F_0 + F_1 + F_2, \quad v = k_2 F_1
\end{aligned} \tag{6.36}$$

The solution is as follows:

$$v = \frac{k_2 SE}{K_M + S + K_I SI} \tag{6.37}$$

Here the maximum reaction rate decreases:

$$v_{max} = \frac{k_2 E}{1 + k_I I} \tag{6.38}$$

In practice the situation is often more complicated.

In nonstationary systems the solution of even simple kinetic problems requires the use of an electronic computer. Numerical solutions give more extensive information about rate constants required in complicated cases.

In fact, even in simple cases the conversion of the substrate to the product involves several steps:

$$E + S \underset{\text{Desorption}}{\overset{\text{Sorption}}{\rightleftharpoons}} ES \underset{\text{Reverse conversion}}{\overset{\text{Conversion}}{\rightleftharpoons}} EP \underset{\text{Sorption}}{\overset{\text{Desorption}}{\rightleftharpoons}} E + P$$

Here E is F_0 and ES $\rightleftharpoons$ EP was earlier denoted by F_1. A study of the dependence of rate constants on temperature allows one to determine the effective values of free energy, enthalpy, and entropy of activation. The reaction has a complicated profile for these quantities (Fig. 6.4). They have been determined for many processes; it has been found that the entropy contribution plays a significant role. For enzymic reactions the quantity $G^{\ddagger}$ usually has values of the order of 40-80 kJ/mole, $H^{\ddagger}$ is of the same order of magnitude, and $S^{\ddagger}$ has values of the order of 40-160 J/(mole·K) deg. Recall that at 300 K the value 140 J/(mole·K) gives a contribution to the free energy equal to 42 kJ/mole. The so-called *compensation effect* has been established. The enthalpy changes ΔH in processes taking place in aqueous solutions are often proportional to changes in entropy ΔS. The same refers to the activation parameters: $H^{\ddagger}$ is proportional to $S^{\ddagger}$. Lumry assumed that this effect is determined by the properties of water: the change in conformation of a protein accompanied by the displacement of water from the active cavity of the globule causes a rearrangement of the surrounding water structure.

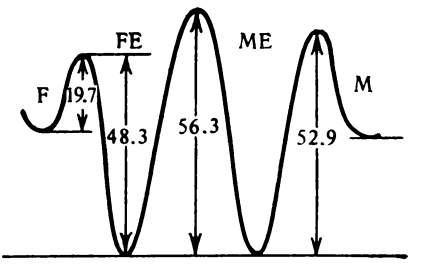


Fig. 6.4. The free-energy curve for the reaction fumarate $\rightleftharpoons$ malate:
F—fumarate; M—malate; FE and ME—their complexes with an enzyme. The numbers are energies in kJ/mole.

The compensation effect may be important to warm-blooded organisms. The smallness of the changes of ΔG and $G^{\ddagger}$ due to compensation means a low sensitivity of enzymic processes to temperature changes in the environment.

The values of $G^{\ddagger}$, $H^{\ddagger}$, and $S^{\ddagger}$ for enzymic reactions are, however, conditional. Kinetic measurements can be made only over a narrow temperature range. The activation parameters strongly depend on the ionic composition of the medium, pH, etc. Here, the Arrhenius equation may not be fulfilled, since the globule undergoes temperature-dependent structural rearrangements. The above interpretation of the compensation effect may prove illusory. These questions have not yet been resolved.

The significant dependence of enzymic activity on the pH of the medium is determined by the electrolytic nature of the enzyme. Enzymic activity, which is characterized in particular by the constant k_2, depends on pH in a bell-shaped manner—k_2 has a maximum at a certain value of pH and falls off with decreasing or increasing pH. A number of theoretical models have been proposed which account for such a dependence in general outline. A true theory, however, requires more detailed information on enzymes than that known to us.

6.3. Chemical Aspects of Enzyme Action

The characteristic features of enzymic catalysis are the high activity and strict specificity with respect to the substrate or a group of substrates.

The molecular activity or *turnover number* of an enzyme is defined as the number of moles of substrate converted per minute by one mole of enzyme at a considerable excess of substrate, $S \gg E$. The properties of enzymes are studied *in vitro*. Many enzymes have been obtained in crystalline form.

Substrates are small molecules or small groups of large molecules. On the contrary, enzymes are macromolecular substances. Hence, the substrate interacts directly with a definite small portion of the enzyme called an *active site*. The nature of the active site, i.e., the amino-acid composition and sequence, and also of the *cofactors* (see page 51) that enter its composition has been established for a number of enzymes. We have already mentioned the enzyme-substrate recognition (page 62). The changes in activity that arise as a result of a chemical modification of a protein make it possible to reveal the functional groups of the active site. Information on its structure is provided by optical and spectral methods and also by the X-ray structural analysis of complexes of enzymes with competitive inhibitors, whose structure is similar to the substrate structure.

The variety of amino-acid residues of the enzyme and of the atomic groups of the cofactor determines the polyfunctionality of the active site—its ability to bind the substrate and catalytic activity. An enzymatic reaction involves many steps and proceeds through a number of chemical conversions. Thus, for example, the active site of *esterases* contains Ser, which undergoes acylation at an intermediate stage of the process. The substrate undergoes the concerted

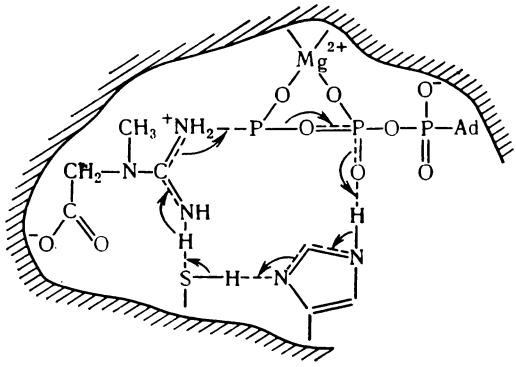


Fig. 6.5. Schematic illustration of the action of creatine kinase. Transition states.

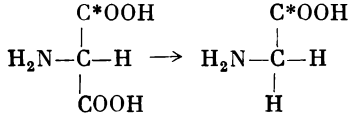
action of the various groups of the active site. As an example may be cited *creatine kinase* which catalyzes the reaction

$$ATP^{4-} + \text{creatine} \rightleftharpoons ADP^{3-} + \text{phosphocreatine}^{2-} + H^{+}$$

Figure 6.5 shows the scheme of the transition state for this reaction. The cofactor is the Mg^{2+} ion. An especially important role is played at the active site by the SH group. The enzyme is inactivated by iodoacetamide and iodoacetate. In the presence of the Mg^{2+} ion and both substrates the SH group is protected from the action of blocking agents. Evidently, the SH group is attached through a hydrogen bond to one of the main groups of the enzyme. The facile transition of a proton in the hydrogen-bonded acid-base system imparts high catalytic activity to this system. The electron transfers in substrates and groups of the active site are shown in Fig. 6.5.

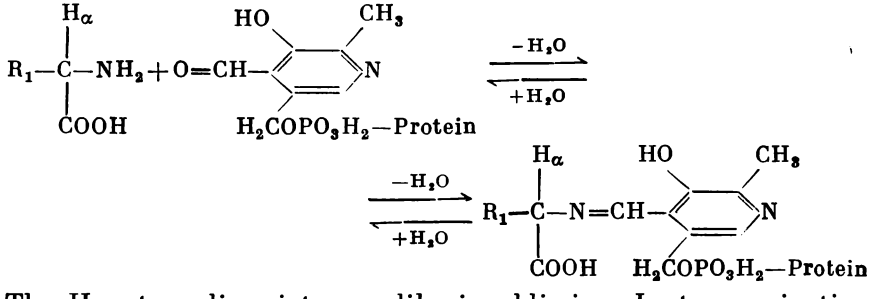
Essential in the formation of enzyme-substrate complexes is the precise mutual orientation of the functional groups of the enzyme and substrate or modifier. The active site is naturally chiral and, thereby, stereospecific. It has been found, with the aid of labeled atoms, that reactions of molecules of the type CAABD occur on the

surface of the enzyme in an asymmetric manner. This refers, for example, to the transformation of aminomalonic acid to glycine:

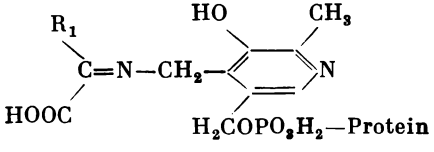


Here C* is the labeled carbon atom. The reaction involves only one COOH group, which is chemically and geometrically indistinguishable from the other group. This is accounted for by the asymmetry of the active site. Since the groups of the enzyme that are bound to the carboxyls are different, the reactions of the carboxyls are also different. Enzymes are capable of discriminating between stereoisomers and the optical antipode of a given substrate is no longer a substrate.

A direct X-ray diffraction study of complexes of lysozyme with inhibiting analogues of substrates—polysaccharides—has shown that the ligand is incorporated into a cavity present in the lysozyme globule and comes into contact with several functional groups of the enzyme (Phillips). The structure of such a complex is depicted in Fig. 6.6. The fact of substrate incorporation has also been established for other systems. For a number of enzymes there has been extensively studied the sequence of chemical conversions, i.e., reaction steps taking place at the active site of the enzyme-substrate complex. For instance, Braunstein and his colleagues have investigated the chemistry of *aspartate aminotransferase* (AAT). This enzyme contains *pyridoxal phosphate* as a cofactor. Pyridoxal phosphate added to a protein reacts with the substrate, an amino acid, chemically, forming *aldimine* (a *Schiff base*):



The H_α atom dissociates readily in aldimine. In transamination reactions, aldimine tautomerizes to ketimine



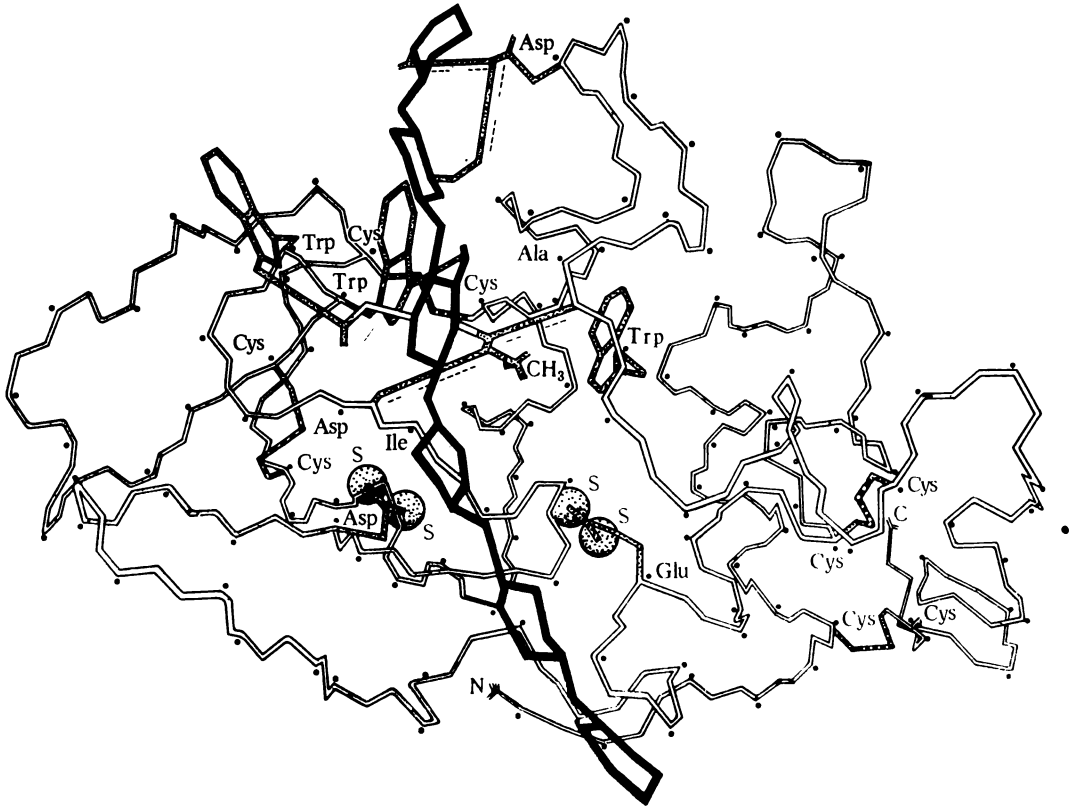
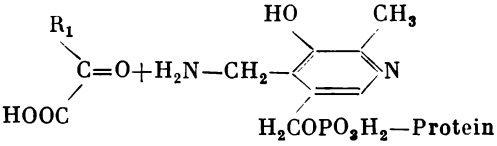


Fig. 6.6. The structure of a complex of lysozyme with β-N-acetalglucosoamine (the latter is black).

which is hydrolyzed by water to give a keto acid:



A pyridoxamine form of the enzyme is formed which reacts with another keto acid containing R_2 in place of R_1, the starting enzyme being regenerated and a new amino acid formed, $R_2CH(COOH)NH_2$. Such is the scheme of *transamination*—the transfer of the amino group.

Practically, all reactions that are catalyzed by enzymes can also take place without enzymes, but at a much slower rate. Therefore, one can model the steps of an enzymatic process in low-molecular-weight "*congruent systems*".

Such a model for a reaction catalyzed by aspartate aminotransferase was constructed by Ivanov and Karpeisky. The first step of the process is the weak binding of an amino acid to the active site. The second step is the nucleophilic addition of the amino group of the substrate to the $N=C\diagup\diagdown$ bond of aldimine. The acceptor of a proton of the NH_2 group is here the negatively charged phenol group of the coenzyme. Thus, the unionized nucleophilic amino group is found to reside near the highly reactive cationic form of the coenzyme. The structure of the aldimine form of the enzyme-substrate complex has been built with the aid of space-filling molecular models. Further, the conversions indicated above take place, some of the steps involving conformational motions—the rotations of the coenzyme ring. The scheme of successive events at the active site of aspartate aminotransferase is shown in Fig. 6.7; the rotations of the coenzyme are marked by a circular arrow. The multipoint binding to the active site is responsible for the stabilization of the active electronic configuration which is thermodynamically unfavourable in solution and for the appropriate orientation of the interacting groups. Each step of the process involves a structural rebuilding that provides the optimum occurrence of the succeeding step. In the language of thermodynamics this means that what happens at the active site is, in fact, the equilibration of the free-energy levels of the intermediate compounds. Indeed, the absorption spectrum of the enzyme-substrate complex shows the absorption bands of all the important intermediate compounds involved in the reaction, whereas the corresponding congruent system displays only the band at 330 nm.

We have discussed these examples in order to show the possibilities of deciphering the interactions between the active site of the enzyme and ligands. Chemistry reveals the behaviour of the

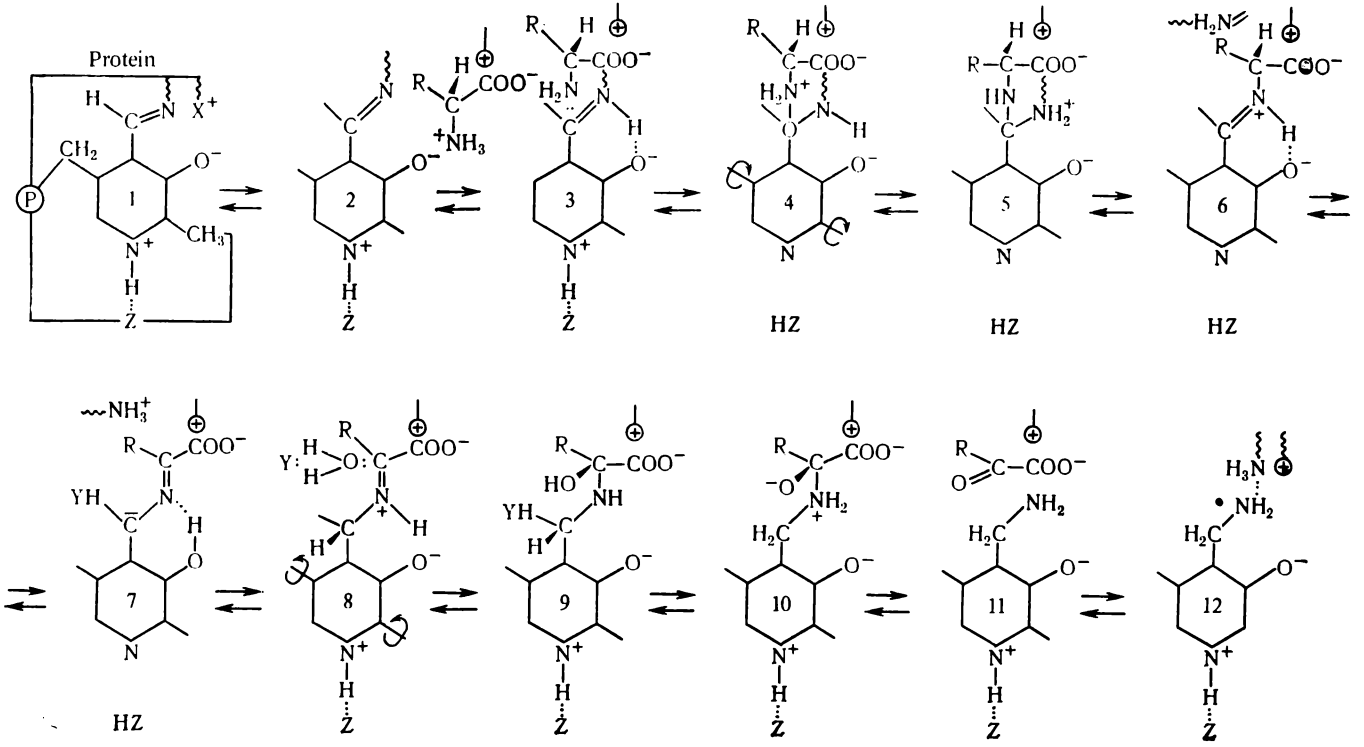


Fig. 6.7. Schematic representation of successive events at the active site of aspartate aminotransferase (AAT).

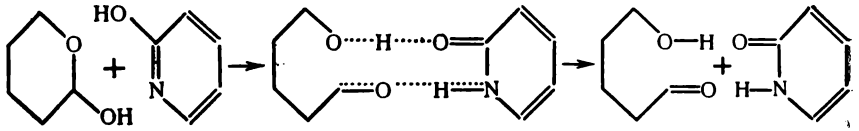
functional groups of enzymes and cofactors. This is, however, insufficient for a quantitative explanation of enzymic activity, which is characterized by a decrease in energy of activation. What is necessary for enzymatic catalysis is the entire protein globule. One cannot cut off a portion of the protein chain without an unfavourable change of the enzyme activity. Chemistry does not provide an answer to the question of the role of the globular structure, being confined only to the description of the events that occur at the active site. These problems are the concern of the physics of enzymes.

Summing up the results of chemical investigations of enzymes, Braunstein pointed out the qualitative factors responsible for their action.

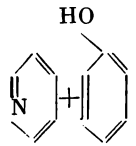
1. A high affinity between enzyme and substrate, i.e., a high probability of formation of an enzyme-substrate complex which is equivalent to a sharp increase in concentration of reactants (*the proximity effect*).

2. A strict mutual orientation of reactants, cofactors, and the active site (*the orientation effect*). In ordinary homogeneous reactions the probability of three or more interacting molecules being strictly oriented with respect to one another is very low.

3. The action on the substrate of nucleophilic and electrophilic groups of the active site (*the effect of concerted general acid-base catalysis*). Let us illustrate this effect by using a model. Hydroxypyridine catalyzes the *mutarotation* of glucose, rupturing its six-membered ring and undergoing a tautomeric change:



An analogous reaction involving a mixture of pyridine and phenol



proceeds at a rate 7000 times slower. The cooperation of acidic and basic groups, their concerted action, sharply accelerates the reaction.

4. The activation of the substrate through the redistribution of electron density under the influence of the active groups of the enzyme (*the polarization effect*).

5. The change of the conformations of the protein and substrate in the enzyme-substrate complex (*the induced-fit effect*).

Physics must be concerned with the quantitative aspect of this picture, indicating at the same time to what extent the effects listed above can be separated.

In conclusion we shall consider data that characterize the rates of enzymatic and analogous nonenzymatic processes (Table 6.1).

Table 6.1. Comparison of Rates of Enzymatic Reactions with Rates of Nonenzymatic Counterparts

Enzyme	Nonenzymatic counterpart	Rate of enzymic process v, s^{-1}	Rate of nonenzymic process v_0, s^{-1}	v/v_0
Lysozyme	Hydrolysis of acetal, base catalysis	5×10^{-1}	3×10^{-9}	2×10^{8}
Chymotrypsin	Hydrolysis of amide, base catalysis	4×10^{-2}	1×10^{-5}	4×10^{3}
β-Amylase	Hydrolysis of acetal, base catalysis	1×10^{3}	3×10^{-9}	3×10^{11}
Fumarase	Hydrogenation of alkene, acid-base catalysis	5×10^{2}	3×10^{-9}	2×10^{11}

Note: The rates of nonenzymatic reactions are multiplied here by a factor corresponding to the proximity effect. For two molecules having the size of the H_2O molecule this factor is equal to 55, i.e., is equal to the molar concentration of water.

6.4. Conformational Properties of Enzymes

In order to understand enzymic activity it is necessary to consider the conformational behaviour of the protein macromolecule. The conformational lability of a protein makes possible its interaction with substrates and other ligands. In some conformations the protein binds the substrate more effectively. Simultaneously, there takes place a selection of the conformations of the substrate. The conformations of the protein and substrate that are selected into the enzyme-substrate complex are those which are *structurally complementary* to each other, thus providing the optimal value of the free energy of interaction. During the formation of an enzyme-substrate complex the conformations of the protein and substrate fit together. The conformational changes involved lead to a structural fit between the enzyme and the substrate.

The conformational effects determine significant changes in entropy during the formation of an enzyme-substrate complex. A decrease in entropy that occurs during the fitting interaction can be compensated for by a decrease in enthalpy. A physical analysis of events of this kind requires that phenomena that occur in the water environment be taken into account.

At the turn of the century Emil Fischer proposed the *lock-and-key model* for treatment of the enzyme-substrate interaction. The enzyme and the substrate possess rigid structures, the active site of the enzyme being structurally complementary to the substrate so that it fits with the substrate in much the same manner as a key fits into a lock. A number of facts contradict this model—the enzyme-substrate interaction is probably not a static one, but is rather a dynamic interaction. Koshland advanced an *induced-fit theory* to account for the specificity of the enzyme-substrate interaction. Let us recall the

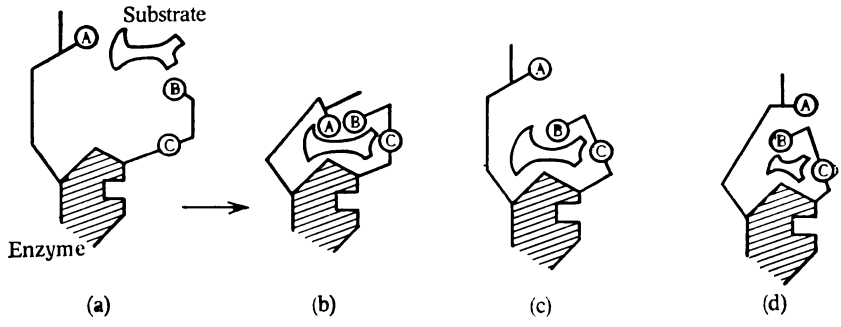


Fig. 6.8. The formation of an enzyme-substrate complex by the induced-fit of enzymic action:
(*a*) enzyme; (*b*) enzyme-substrate complex (ESC); (*c*) and (*d*) complexes with substrate analogues.

basic postulates of this theory, which was designed primarily in an attempt to explain the specificity of enzymes that catalyze bond-transfer reactions:

$$B{-}X + Y \rightleftharpoons B{-}Y + X$$

1. The substrate causes changes in the geometry of the enzyme as it fits into the active site.
2. A delicate orientation of catalytic groups is required for enzyme action.
3. The substrate induces this proper orientation by the change it causes in the geometry of the enzyme.

Of course, the induced-fit theory covers also those cases which involve the alteration of the conformation of the substrate rather than the enzyme or of the conformations of both. The main idea of the theory is the determinative role of the conformational lability of the interacting molecules for enzymatic catalysis.

The three postulates of the theory are illustrated by Fig. 6.8. Figures 6.8*c* and *d* show why molecules structurally similar to the

substrate but different from it in size do not react: the functional groups of the enzyme, A and B, cannot in these cases arrange themselves in a proper manner. According to the Koshland model theory, the substrate specificity and catalytic efficiency of the enzyme are closely related but their mechanisms are different and the reaction can be accomplished only with a definite disposition of the contact and catalytic groups of the enzyme with respect to the substrate molecule. Koshland illustrates this by the example of β-*amylase*. This enzyme acts on the terminal groups of amylose but not on the

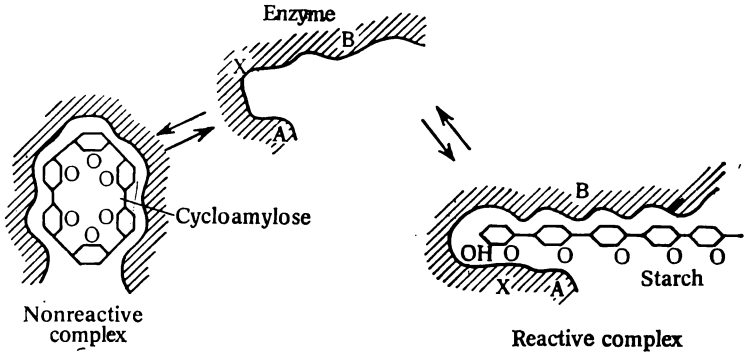


Fig. 6.9. Diagram illustrating the action of β-amylase.

other glucoside linkages of the polysaccharide. Cycloamyloses are the competitive inhibitors of the enzyme. What has just been said is explained by the scheme presented in Fig. 6.9. The reaction occurs only with the contact and catalytic groups of the enzyme, A, B, and X, being spatially arranged in a strictly definite way.

A number of facts really serve as an indication of the conformational changes of enzymes upon their interaction with substrates. In the presence of substrates some enzymes become more rigid, others, on the contrary, are made more labile—they are denatured more easily upon heating. Substrates induce the dissociation of *glutamate dehydrogenase* and *hexokinase* into subunits. The substrate affects the reactivity of the amino-acid residues of the enzyme. The absorption spectrum of *chymotrypsin* is changed when it interacts with the substrate and these changes may be treated as being caused by the change in conformation. The changes in conformation are seen in the luminescence spectra of both aromatic amino-acid residues and dyes sorbed on the protein. Spectropolarimetric methods have revealed changes in α-helicity that arise upon interaction

of enzymes with substrates, coenzymes, and other ligands. Information on conformational changes in enzyme-substrate complexes is also furnished by the EPR spectra of enzymes containing paramagnetic labels, NMR spectra, etc.

The X-ray diffraction studies of enzyme-substrate complexes yield direct information about conformational transformations. Upon incorporation of the substrate into *lysozyme* (page 201) the crevice contracts and "squeezes" the substrate more tightly. The displacements of the amino-acid residues of the protein are not great but they

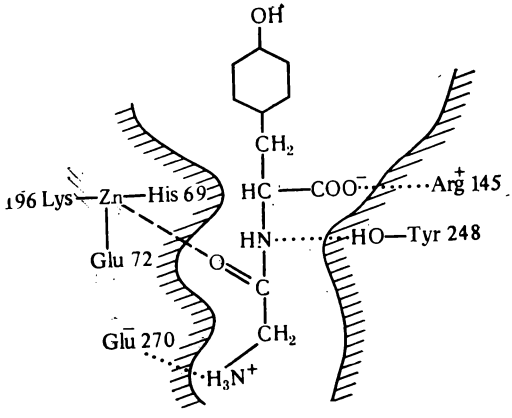


Fig. 6.10. The active site of carboxypeptidase.

are appreciable—the residue Try 62 is shifted by 0.075 nm. This is accompanied by a change in conformation of the substrate analogue—a slight rotation of the carbohydrate rings about the glycoside linkage.

The presence of a crevice designed specifically for the substrate has also been detected for other enzymes. This reveals groups of the active site. A direct confirmation of the induced-fit concept has been obtained in an investigation of *carboxypeptidase*. In the presence of the substrate glycyltyrosine the amino-acid positions are changed significantly. The active site is a deep cavity similar to the mouth of a spider with its "legs" ready to direct the substrate to the cofactor—to the Zn atom. One "leg" that contains Tyr 248 is directed to the NH group of the substrate, the Arg 145 of the second interacts with the carboxyl group of the substrate, and the Glu 270 of the third reacts with the terminal amino group (Fig. 6.10).

As has been pointed out earlier (page 146), X-ray diffraction analysis has recently made it possible to determine the relative mobility (conformational and vibrational) of various amino-acid

residues in proteins. The investigations of myoglobin and lysozyme have demonstrated that the residues within the globule are much less mobile than those residing on its surface.

Thus, in enzyme-substrate complexes there is achieved a structural fit realized in the cavity of the enzyme molecule. As shown by a comparison of all the structures studied by X-ray diffraction, their common specific feature is that the inner surface of the cavity is predominantly formed by nonpolar residues. As a result of hydrophobic interactions, the polar residues are moved to the surface. The nonpolar interior of the protein molecule has a low dielectric constant, which facilitates the occurrence of electric interactions. The enzyme is not only a specific reagent but serves also as a medium for the reaction. As Perutz remarks, we may ask ourselves why chemical reactions that normally require powerful organic solvents or strong acids and bases can proceed in aqueous solution near neutral pH in the presence of enzymic catalysts. Organic solvents have advantages over water, providing a medium of low dielectric constant in which intensive electric interactions between the reagents may occur. The nonpolar interior of the enzyme provides a living cell with equivalents of organic solvents that are used by chemists.

These are all qualitative conceptual models, requiring no physical theory of enzymic catalysis. The paths leading to such a theory (not yet constructed) will be described further in the text.

6.5. Physics of Enzyme-Substrate Interactions

The physical specificity of enzymic catalysis is primarily due to the role played by the globule, whose size is an order of magnitude larger than the size of the substrate. Naturally, a question arises as to the causes for the high efficiency of such a system in the course of a chemical conversion in aqueous solution at a normal pressure and physiological temperature.

Upon binding to the active site the substrate molecule is transferred from an aqueous environment to an environment produced by amino-acid residues. The substrate is found to be surrounded by a medium of low dielectric constant, in which strong electric interactions may occur between the reactants and the polar groups of the enzyme. Elaborating on the enzyme-solvent concept, Perutz arrives at the conclusion that electrostatic interactions make a major contribution to the energetics of enzymic catalysis, i.e., to the reduction of the energy of activation caused by the enzyme. The difference between the enzyme and the aqueous solution consists in that the active site of the enzyme accommodates dipoles with a fixed orientation with respect to the charged groups of the substrate, even if the field of these charges is small. As a result of this orientation, enzymes are capable of stabilizing ion pairs and other charge distributions to a

considerably greater extent than water. In water solutions the electrostatic attraction is not strong, since any force created by the approach of opposite charges is counterbalanced by the reorientation of the solvent dipoles. In enzymes such reorientation is impossible. The quantitative calculations carried out on this basis for lysozyme have enabled an estimate of the decrease in the activation energy of a catalyzed reaction in accordance with experiment.

Undoubtedly, electrostatic interactions are of importance for enzymic catalysis, especially in the case of ionic reactions. Lysozyme is in this sense a specific enzyme and the calculations made cannot be regarded as being of universal value.

The protein globule is a dynamic system. There have been made a number of interesting attempts at a physical interpretation of the behaviour of such a system upon its interaction with the substrate.

Obviously, the energy that can be expended by the enzyme in order to speed up the reaction (i.e., to lower effectively the activation energy barrier) can come from only one source—it is a part of the free energy liberated upon binding (sorption) of the substrate to the enzyme. The assumption of the accumulation of the heat energy of the environment in the enzyme and of its utilization in the reaction would mean a so-called perpetual-motion machine of the second kind. Thus, the energy is liberated upon sorption of the substrate. A hypothesis has been suggested, according to which this energy is transformed to the energy of the elastic vibrations of the globule, which behaves like a liquid drop. The frequencies of such vibrations lie in the hypersonic region—up to 10^{13} s^{-1}. The standing waves in the liquid drop can create a vibration maximum in the region of the active site and the energy of elastic vibrations can activate the substrate molecule. Quantitative estimates based on this idea have shown that the energy of elastic vibrations of the globule may indeed amount to 20-40 kJ/mole and provide a considerable decrease in the effective activation barrier.

This hypothesis is attractive but has not been proved so far. It is also not clear at what rate the elastic energy of the globule is dissipated into the environment.

We shall nevertheless make a simple estimate of the acceleration of the reaction based on the assumption of the transformation of the sorption energy to the energy of the enzyme-substrate complex. Suppose that the structural fit between enzyme and substrate in the enzyme-substrate complex brings both the protein and the small molecule into a stretched state. Let the length of the unstretched substrate molecule be l_0 and the length of the enzyme cavity into which the substrate fits, l. The changes in length of the substrate molecule and the enzyme are, respectively, equal to x and y. Then, $x + y = l - l_0$ and the condition for the equality of the elastic

forces has the form

$$k_S x = k_E y \tag{6.39}$$

where k_S and k_E are the coefficients of elasticity of the substrate and enzyme. We find

$$x = \frac{l - l_0}{1 + k_S/k_E} \quad \text{and} \quad y = \frac{k_S}{k_E} \frac{l - l_0}{1 + k_S/k_E} \tag{6.40}$$

The elastic energy of the substrate, which determines the decrease in the activation energy, is given by

$$\Delta E = \frac{1}{2} k_S x^2 = \frac{1}{2} k_S \left(\frac{l - l_0}{1 + k_S/k_E} \right)^2 \tag{6.41}$$

The order of magnitude of k_E corresponds to the product of the linear size of the globule by the modulus of elasticity $L\varepsilon$. For proteins $L \approx 5$ nm and ε is about 10^3 J·cm^{-3}. Hence, $k_E \approx 5 \times 10^{-2}$ N/cm. The largest amount of energy of elastic deformation is concentrated at the weakest point of the substrate molecule. The deformation of valence angles occurs more readily than that of valence bonds. At the same time, the energy stored at the angular degrees of freedom of the molecule can be transferred to a valence bond and can lower the activation energy of its rupture. The elastic modulus k_S, which corresponds to low-frequency deformation vibrations ($\nu \sim 10^{13}$ s^{-1}) is approximately equal to 0.15 N/cm^{-1}. We now assume that $\Delta E = 31.5$ kJ/mole (when the energy of activation is diminished by this amount, the reaction rate increases by 10^5 times). Then, $x \approx 0.08$ nm, $y \approx 0.23$ nm, and the elastic energy of the enzyme $\frac{1}{2} k_E y^2 \approx 88$ kJ/mole. This means that the total energy expended during the sorption on elastic deformation is 171 kJ/mole. This value is not too great, considering that the sorption occurs as a result of multipoint binding, i.e., the formation of many chemical and weak bonds between enzyme and substrate. The sorption energy observed is the difference between the true sorption energy and the energy of elastic deformation of the enzyme and substrate.

The stretched state model (the so-called "rack" model) considered above is a static one. If we assume that the elastic system is a dynamic one and there arises a resonance of the vibrations of the substrate and enzyme molecules, then the same acceleration will be attained by an average elastic energy four times lower than in the static case, since pulsations periodically double the vibrational amplitude.

This spectacular model deals with the vibrations of atomic nuclei that arise as a result of the formation of an enzyme-substrate complex. This interaction is accompanied by changes in the states of the electron shells of the substrate and atomic groups of the active site. The electron shells experience a perturbation due to interactions

in the enzyme-substrate complex. The conversion of substrates to products is a chemical process, i.e., a change in the states of the electron shells of molecules. Just as in any other chemical reaction, this involves transfers of atomic nuclei. Among the motions of atomic nuclei, the lowest energy is expended on low-frequency deformation vibrations and rotations about single bonds, i.e., conformational changes. The conformational changes in the enzyme-substrate complex were discussed in Sec. 6.4. Of prime importance for enzymatic catalysis are the interactions of electronic and conformational degrees of freedom—*electronic-conformational interactions* (ECI). Such interactions have been treated in the works of Volkenshtein, Blumenfeld, Chernavsky, Dogonadze, and their coworkers.

6.6. Electronic-Conformational Interactions

In order to understand the nature of electronic-conformational interactions it is helpful to make use of a spectacular model of the interaction between electrons and atomic nuclei—electrons in a potential box with infinitely high mobile walls. A set of atomic nuclei

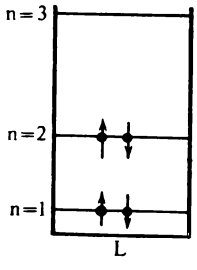


Fig. 6.11. Electrons in a potential box.

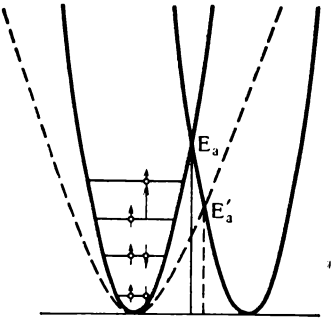


Fig. 6.12. The lowering of the activation barrier caused by an electron-conformation interaction.

of a molecule or its functional part is modelled by means of such a box (Fig. 6.11). Electrons are placed in the box, $2n$ electrons occupy n levels. The possible values of the energy of the electrons inside the box are easily calculated on the basis of the concepts of de Broglie standing waves with nodes at the walls. If the width of the box is L, the standing waves have wavelengths equal to

$$\lambda = 2L/n, \quad n = 1, 2, 3, \ldots \tag{6.42}$$

The velocity of an electron in the box is found from the de Broglie relation (m_e is the electron mass):

$$v = \frac{h}{m_e \lambda} = \frac{nh}{2m_e L} \tag{6.43}$$

Inside the box the energy of the electron is equal to its kinetic energy:

$$E = \frac{m_e v^2}{2} = \frac{n^2 h^2}{8m_e L^2} \tag{6.44}$$

The electron exerts a pressure on the wall which is given by

$$f = \left| \frac{dE}{dL} \right| = \frac{n^2 h^2}{4m_e L^3} \tag{6.45}$$

At equilibrium these forces are compensated for by the interactions with the surroundings of the box. A change in equilibrium results either from the excitation of electrons in the system or from the addition of electrons. Either of these leads to an increase in pressure. Being no longer at equilibrium, the walls of the box move and pass to a new equilibrium position at an increased distance from each other, $L + \Delta L$. In other words, the work of displacement of the nuclei, $f\Delta L$, is done at the expense of the energy of electrons—according to Eq. (6.44) the electronic energy decreases with increasing L.

If we consider not a rectangular potential well but a parabolic one, i.e., if we make use of the electron-harmonic oscillator model, it will be easy to show how the electronic-conformational interactions lower the activation barrier (Fig. 6.12). The expansion of the parabola brought about by the added pressure force means a decrease in the elastic coefficient for the oscillator. The point of intersection with the second parabola which corresponds to the final state of the enzyme-substrate complex is shifted and its ordinate, which is the energy of activation, is decreased. One can show that for the energy of activation to be considerably diminished, i.e., for the reaction to be greatly accelerated, a low efficiency of the system is sufficient.

The above treatment is only of illustrative value. It shows, however, that the energy of electronic excitation is converted to the work of displacement of atomic nuclei, i.e., to the conformational energy. As a result, the reaction is accelerated.

Lumry and Biltonen assumed that the profile of the conformational free energy along the reaction coordinate is complementary to the profile of the chemical (electronic) free energy (Fig. 6.13). As a result, the activation barriers are lowered in the overall profile.

More rigorous quantum-mechanical approaches to electronic-conformational interactions have been worked out by Dogonadze and his coworkers. The solution of the problem, i.e., the calculation of the probability of the substrate being converted to the product, requires finding the potential energy surface of the electronic states

of the system, their dependences on the positions of atomic nuclei. For nonadiabatic electron-and-atom transfer reactions, i.e., for reactions accompanied by conformational shifts of nuclei, the key role is played by the *Franck-Condon principle.* Since the nuclei move much more slowly than electrons, the probability of the reaction is different from zero provided that the electron levels of the initial and final states are equilibrated due to conformational motions. Figure 6.14 presents the corresponding diagram (Q is the conformational coordinate). The electronic rearrangement is preceded by a complicated motion from Q_{0i} to Q^* on the potential-energy surface. After the electron rearrangement in the region of intersection of the surfaces the system passes to a conformationally excited state, which strongly differs from the equilibrium state, and relaxes then to a new equilibrium state; the $Q^* \rightarrow Q_{0f}$ transition is realized. The kinetics of the process is governed by electron-conformation interactions.

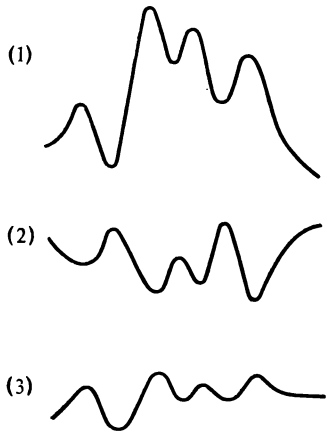


Fig. 6.13. The complementarity effect: (*a*) the profile of the chemical free energy; (*b*) the profile of the conformational free energy; and (c) the overall profile.

In a theoretical calculation the parameters of the system—the enzyme-substrate complex and the solvent that surrounds it—may be represented by the sum of the harmonic oscillations of the nuclei. What one finds is the energy of the nuclear (i.e., conformational) subsystem in its initial and final states. Conformational oscillators are slow, their energy being much less than thermal energy, $\hbar\omega \ll kT$. They are classical oscillators, in contrast to quantum-mechanical, electronic ones. Theory leads to the expression of the probability of the transition per unit time between the initial and final states. According to Arrhenius, this expression depends exponentially on the sum of the free energy of the reaction and the energy of reorganization of the conformational subsystem and the solvent. The meaning of the energy of conformational reorganization is disclosed in Fig. 6.15 (Q_{0i} and Q_{0f} are the classical equilibrium coordinates of the initial and final states of the system; E_a is the energy of activation; E_r is the energy of conformational reorganization; ΔI is the reaction heat). Thus, the system is moving along a series of classical coordinates and reaches the region of intersection of potential surfaces with the probability indicated. The time during which the system resides

in the transition region is determined by the effective frequency of motion and the reorganization energy. It is during this time period that the system passes from the initial to the final state.

The theory briefly discussed above has a distinct and strict physical content. However, quantitative computations based on this theory are very complicated and require a knowledge of many parameters. Therefore, a comparison of the theory with experiment is a difficult task.

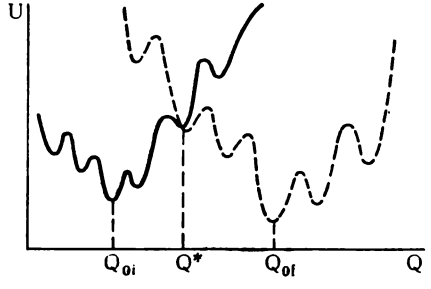


Fig. 6.14. Diagram showing an electronic-conformational transition.

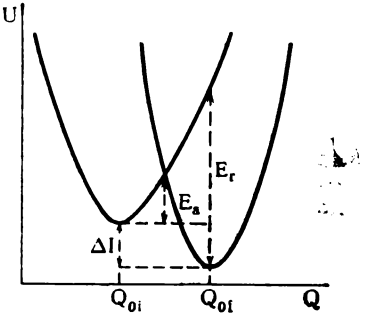


Fig. 6.15. The parameters of an electronic-conformational transition.

The overall result of the electronic-conformational interactions consists in that the electron rearrangement during the course of an enzymatic reaction is associated with the change in conformation of the protein. This situation can be described in the language of solid-state physics. The displacement of an electron or electron density in a macromolecule causes a conformational rearrangement. It may be treated as the excitation of longwave phonons and the system electron plus the deformation of the macromolecule is found to be similar to the polaron. We shall call such a system the *conformon.*

In contrast to the polaron, the conformon is not a true quasiparticle capable of moving without dissipating energy at large distances. Rapid dissipation is determined by the inhomogeneity and aperiodicity of the globule. However, for an enzymatic process to be accomplished it is sufficient that a conformational change be effected within a few peptide linkages. The subsequent conformational rebuilding of the globule will depend not on electronic-conformational interactions but on the cooperativity of the macromolecule.

Experimental information on electronic-conformational interactions can be obtained by studying changes in chemical (electronic)

properties of biopolymers caused by the change of the ligand or cofactor and by observing the attendant changes in conformational properties of the macromolecule as a whole. An investigation of this kind was carried out on aspartate aminotransferase-ligand (cofactor) systems. What was studied is the denaturation in a urea solution of an apoenzyme, normal choloenzyme, and a number of complexes of the apoenzyme with various chemical modifications of the coenzyme. The denaturation isotherms were found to be strongly dependent on the nature of the coenzyme. This means that the electronic changes in the active site have a strong effect on the conformational properties of the protein as a whole. A quantitative analysis of the results obtained and a comparison of an enzymatic process catalyzed by aspartate aminotransferase with the reaction in a congruent model system have shown that the energy levels of various intermediate forms are equilibrated in the enzyme system and that the activation barriers are lowered. This is a manifestation of the complementarity of the profiles of chemical (electronic) and conformational free energies (page 213). The complementarity is determined by electronic-conformational interactions.

The problem of electronic-conformational interactions is effectively tackled by means of the qualitative methods of quantum chemistry which allow one to clearly visualize the electron shell of the substrate, which is influenced by the perturbing action of the active-site groups. Of particular importance in the physics of enzymes at present are well-substantiated qualitative models that enable one to get an insight into the phenomena observed and to evaluate the order of magnitude of the values of the characteristic parameters (the energy of activation, etc.).

The important objects for the study of electronic-conformational interactions are *metalloenzymes*, in particular enzymes whose cofactors are ions of *transition metals*. Such ions serve as bright labels; their behaviour, particularly the change of the electronic state, can be easily studied by using the methods of optical spectroscopy, EPR, etc. At the same time, the electronic state of the metal in a metalloenzyme exerts a strong influence on the conformational properties of the protein. In its turn, the state of the metal depends on the protein environment. Vallee has revealed appreciable differences in electronic properties of transition metals between enzymes and low-molecular-weight complex compounds. The metal ion in the protein is in a stressed, "entatic" state. Here we enter the field of bioinorganic chemistry. The properties of some metal-containing proteins (hemoglobin and myoglobin) will be considered at a later time (see Sec. 6.8); we shall also discuss electron-transfer processes in metalloenzymes of the respiratory chain—in *cytochromes*. In Sec. 13.4 we shall again return to the consideration of electronic-conformational interactions.

6.7. Cooperative Properties of Enzymes

The steady-state kinetics of enzymic reactions often differs from Michaelis-Menten kinetics (page 194). The curve showing the dependence of reaction velocity v on substrate concentration S and/or modifier concentration displays inflections, maxima, and plateaus. An example is given in Fig. 6.16, which presents the curve of v versus S

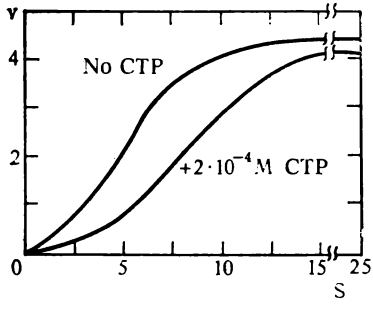


Fig. 6.16. The curve of v vs. S for aspartate transcarbamylase. The abscissa shows the concentration of aspartate in 10^{-3} mole/litre.

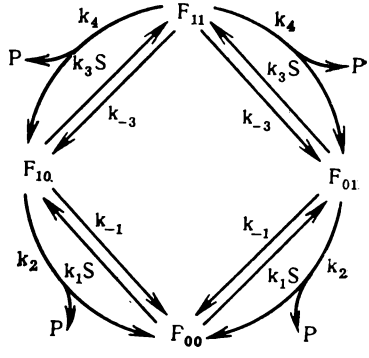


Fig. 6.17. Schematic representation of the reaction with an enzyme consisting of two subunits.

for *aspartate aminotranscarbamylase* (S is aspartate) in the absence and presence of a modifying reagent—cytidine triphosphate (CTP). In the first case, the curve is sigmoid rather than hyperbolic in shape.

As a rule, these features are accounted for by the presence of a quaternary structure in the enzyme and by the interaction of subunits. Thus, the behaviour of the enzyme is cooperative—the affinity for the substrate and the catalytic activity of a particular subunit (protomer) depend on what states the other subunits are in—whether they have bound the substrate or not.

Let us consider a simple model of the enzyme molecule composed of two interacting identical subunits. Each of the subunits has an active site. The scheme of the steady-state process is depicted in Fig. 6.17. The enzyme molecule can exist in three states: F_{00} (both sites are free), $F_{01} = F_{10}$ (one site is occupied by the substrate and the other is free), and F_{11} (both sites are occupied by the substrate).

The equations for steady-state kinetics have the form (cf. Fig. 6.17):

$$\begin{aligned}
\dot{F}_{00} &= -2k_1SF_{00} + 2(k_{-1} + k_2)F_{10} = 0 \\
\dot{F}_{10} &= 2k_1SF_{00} - 2(k_{-1} + k_2 + k_3S)F_{10} + 2(k_{-3} + k_4)F_{11} = 0 \\
\dot{F}_{11} &= 2k_3SF_{10} - 2(k_{-3} + k_4)F_{11} = 0
\end{aligned} \tag{6.46}$$

Here

$$F_{00} + 2F_{10} + F_{11} = E = \text{const} \tag{6.47}$$

The rate of formation of the product is given, in accordance with Fig. 6.17, by

$$v = 2k_2F_{10} + 2k_4F_{11} \tag{6.48}$$

Using Eqs. (6.46) and (6.47), we find

$$v = 2k_2ES\frac{\alpha S + \beta K}{S^2 + 2\beta KS + \beta K^2} \tag{6.49}$$

where $\alpha = k_4/k_2$, $\beta = K'/K$, $K = (k_{-1} + k_2)/k_1$, $K' = (k_{-3} + k_4)/k_3$.

Cooperativity, i.e., the interaction of two active sites, is expressed in terms of the difference of α and/or β from unity. Indeed, on the condition $\alpha = \beta = 1$

$$v = 2k_2ES\frac{K + S}{(K + S)^2} = \frac{v_{max}S}{K + S} \tag{6.50}$$

that is, the kinetics coincides with Michaelis-Menten kinetics. On the contrary, if this condition is not satisfied, then the curve of v versus S is likely to show a point of inflection or a maximum or both. Analysis of formula (6.49) shows that points of inflection can appear with the following conditions:

at $\alpha < 0.5$ and any β

at $1 > \alpha > 0.5$ and $\alpha/(2\alpha - 1) > \beta > \alpha^2/(2\alpha - 1)$

at $\alpha > 1$ and $\alpha^2/(2\alpha - 1) > \beta > \alpha/(2\alpha - 1)$

A maximum is possible only at $\alpha < 0.5$ and any β.

The model described above expresses *direct cooperativity*—the rate constants for the states of the system in which the substrate occupies one or two active sites are different.

The equilibrium function $\bar{Y}$ of saturation of enzyme with substrate for the two-site model is expressed as follows:

$$\bar{Y} = \frac{2F_{10} + 2F_{11}}{2(F_{00} + 2F_{10} + F_{11})} = \frac{S^2 + \beta KS}{S^2 + 2\beta KS + \beta K^2} \tag{6.51}$$

The function $\bar{Y}(S)$ changes to the Langmuir adsorption isotherm at $\beta = 1$. At $\beta \neq 1$ the $\bar{Y}(S)$ curve may have a point of inflection but not a maximum.

Thus, the fact that the v (S) and $\overline{Y}$ (S) curves show some singularities under steady-state (accordingly, equilibrium) conditions is an indication of cooperativity. The absence of such singularities does not necessarily mean that no cooperativity exists.

Another model—the *indirect cooperativity model*—has been proposed by Monod, Wyman, and Changeux (it will be referred to as *the M-W-C model*). The protein molecule is modelled by an oligomer consisting of two or more identical subunits—protomers which occupy equivalent spatial positions. Thus, the molecule has some symmetry. It may be built up in an isologous or heterologous manner; in the latter case, the oligomer may be long in length (Fig. 6.18). Each subunit contains one binding site for each ligand (substrate or modifier).

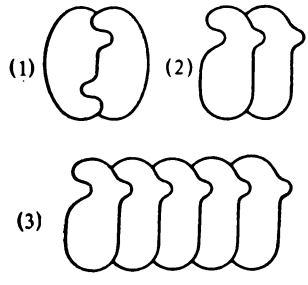


Fig. 6.18. The possible structures of isologous (*1*) and heterologous (*2*) dimers and of a heterologous polymer (*3*).

It is presumed that the oligomer as a whole can exist in two (or more) conformational states. The affinity for the ligand varies with a change in state of the oligomer. Such a system is cooperative. Without loss of generality let us consider a dimer, which can exist in two states designated R and T. In either of these states the dimer is capable of addition of 0, 1 or 2 molecules of ligand S. Thus, the number of different states of the dimer is 6: R_{00}, $R_{10} = R_{01}$, R_{11}, T_{00}, $T_{10} = T_{01}$, T_{11}. The equilibrium conditions are as follows:

$$T_{00} = LR_{00}$$

$$R_{10} = 2R_{00}S/K_R, \quad R_{11} = \frac{1}{2} R_{10}S/K_R = R_{00}S^2/K_R^2 \qquad (6.52)$$

$$T_{10} = 2T_{00}S/K_T, \quad T_{11} = \frac{1}{2} T_{10}S/K_T = T_{00}S^2/K_T^2$$

Here K_R and K_T are the dissociation constants for S in states R and T, respectively; L is the equilibrium constant for the $R \rightleftharpoons T$ transition in the absence of S. The total concentration of enzyme is constant:

$$R_{00} + R_{10} + R_{11} + T_{00} + T_{10} + T_{11} = E \qquad (6.53)$$

The substrate saturation function for the enzyme has the following form:

$$\overline{Y} = \frac{R_{10}+2R_{11}+T_{10}+2T_{11}}{2(R_{00}+R_{10}+R_{11}+T_{00}+T_{10}+T_{11})} = \\ = \frac{K_R(1+Lg)(1+Lg^2)^{-1}S+S^2}{K_R^2(1+Lg)(1+Lg^2)^{-1}+2K_R(1+Lg)(1+Lg^2)^{-1}S+S^2} \quad (6.54)$$

where $g = K_R/K_T$. The $\overline{Y}$ vs. S plot has a point of inflection. At $g = 1$ or with L tending to zero or infinity, cooperativity disappears and the last expression again transforms to the expression for the Langmuir isotherm:

$$\overline{Y} = \frac{S}{K_R+S} \quad (6.55)$$

The corresponding expression for the rate of conversion of the substrate, the one obtained from the equilibrium condition, has the form

$$v = 2Ek\frac{1+\varkappa Lg^2}{1+Lg^2}\frac{K_R(1+\varkappa Lg)(1+\varkappa Lg^2)^{-1}S+S^2}{K_R^2(1+Lg)(1+Lg^2)^{-1}+2K_R(1+Lg)(1+Lg^2)^{-1}S+S^2} \quad (6.56)$$

where k is the rate constant for states R_{10} and R_{10}, and $\varkappa k$ is the rate constant for the T_{10} and T_{11} states.

The cooperativity in the M-W-C model is indirect, since the addition of the ligand causes a shift in the conformational equilibrium. If we extend Eq. (6.54) to a system composed of n protomers, we obtain:

$$\overline{Y} = \frac{Lgx(1+gx)^{n-1}+x(1+x)^{n-1}}{L(1+gx)^n+(1+x)^n} \quad (6.57)$$

where $x = S/K_R$.

At $g \ll 1$ and not too large L and $\varkappa$ formulas (6.54) and (6.56) assume the following form for the dimer:

$$\overline{Y} \approx x\frac{1+x}{L+(1+x)^2} \quad (6.58)$$

$$v \approx 2kx\frac{1+x}{L+(1+x)^2} \quad (6.59)$$

Let us assume that, apart from the substrate S, the dimer is also acted on by an inhibitor I and an activator A and that each of the two protomers contains three active sites—for S, I and A. Assuming for simplicity that the dimer adds on I only in the T-state and A only in the R-state, we obtain at $g \ll 1$ the substrate saturation function for the enzyme:

$$\overline{Y} \approx x\frac{1+x}{L'+(1+x)^2} \quad (6.60)$$

where $L' = L(1 + I/K_I)^2 (1 + A/K_A)^{-2}$. The inhibitor enhances indirect cooperativity, whereas the activator diminishes it.

If cooperativity is positive, i.e., the addition of a ligand activates the enzyme, the models of direct and indirect cooperativity may lead to equivalent results. However, in contrast to the M-W-C model, the direct cooperativity model can also describe negative

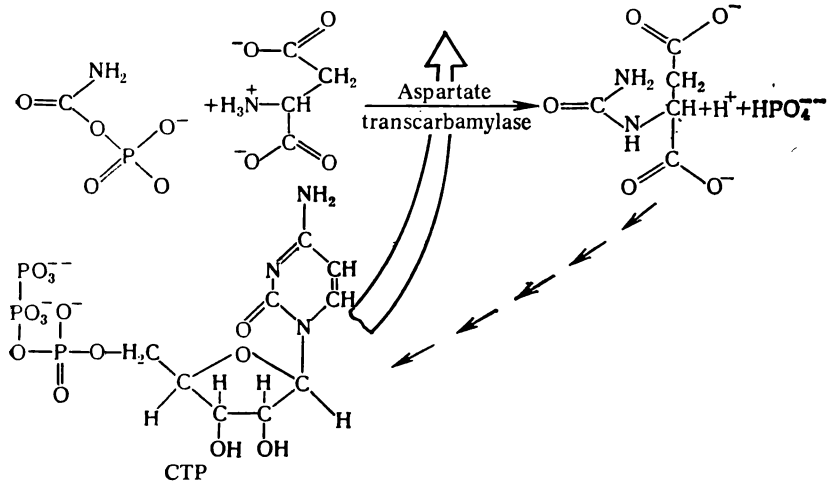


Fig. 6.19. The allosteric feedback.

cooperativity, a phenomenon in which the affinity towards the corresponding ligand is reduced as the active sites are saturated more and more (Koshland).

Umbarger detected in 1956 the existence of sequential enzymatic reactions, in which the end metabolite affects the activity of the enzyme that catalyzes the first reaction in the sequence. We know at present many such systems, called *allosteric* systems. In particular, cases of inhibition have been established, whose kinetics is similar to the kinetics of competitive inhibition, though the structure of an allosteric inhibitor is different from that of the substrate. For instance, aspartate transcarbamylase is an enzyme that catalyzes the first step in the sequence of conversions leading to the synthesis of CTP. This end product of the chain of reactions inhibits aspartate transcarbamylase (Fig. 6.19). The expediency of such a system is obvious. If the end metabolite is immediately consumed in subsequent reactions, then the entire system works. However, if an excess of the end metabolite appears, it stops the operation of the entire system.

Allosterism is an important phenomenon that provides for the regulation of enzymic action at a molecular level. The presence of the regulation of biochemical processes means the existence of direct and feedback channels, through which appropriate information is transmitted. The transmission of information in the cell is the transmission of chemical signals, i.e., molecules and ions. The cell contains substances whose reactivity is determined by the action of molecular signals on them. Since such signals come from enzymes, it is enzymes that are subject to the effect of chemical feedback. One may come to the logical conclusion of the existence of regulatory enzymic systems. We have already said that an enzyme may be regarded as a signal transformer: the input signal, i.e., the substrate, is transformed to the output signal, i.e., the product. If the output signal affects the operation of the transformer, a feedback is realized, positive or negative, depending on whether the end product activates or inhibits the function of the enzyme.

An investigation of allosteric enzymes has shown that they possess a quaternary structure and cooperativity. This is evidenced by the v vs. S curve for aspartate transcarbamylase (see Fig. 6.16). Cytidine triphosphate (CTP), in the presence of which the sigmoidicity of the curve disappears, is an allosteric inhibitor. Another example of allosteric enzymes is threonine deaminase. The kinetics of the conversion of threonine by this enzyme may be described by the formula

$$\log \frac{v}{v_{max} - v} = n \log S - \log K \qquad (6.61)$$

and the kinetics of allosteric inhibition is depicted as follows:

$$\log \frac{v}{v_0 - v} = \log K' - n' \log I \qquad (6.62)$$

Here v is the steady-state velocity; v_{max} is the maximum velocity; v_0 is the velocity at $I = 0$; K and K' are constants. For Michaelis-Menten kinetics the following condition must be fulfilled: $n = n' = 1$. Changeux found that $n = 1.37$ and $n' = 1.86$. These results are easily explained by cooperativity—the interaction of subunits. Similar phenomena will be discussed in the next section for hemoglobin. It has been established that each subunit of allosteric enzyme has one active site for the substrate and the other for the allosteric effector. More complicated cases are also observed.

The fact that allosteric enzymes exhibit a quaternary structure determines the possibility of their dissociation into subunits under the influence of substrates and allosteric effectors. There takes place a shift of the equilibrium between the oligomeric forms of an allosteric enzyme. This is also manifested in the kinetics.

The singularities on the v vs. S and v vs. I plots may also arise in the absence of cooperative interactions due to the nonequilibrium

conformational properties of the enzyme. Let us assume that the enzyme molecule that has converted the substrate into the product emerges from the reaction in an active conformational state. If the relaxation time, i.e., the time of return to the original unperturbed state, is longer than the time between the encounters of the enzyme with the substrate or of the same order, then the kinetics may simulate cooperative kinetics. The scheme of such a process is shown in Fig. 6.20. Here F_0 is the substrate-free enzyme molecule in the original conformation; F_1 is the inactive enzyme-substrate complex; F_2 is an active enzyme-substrate complex; F_3 is the free enzyme in an active conformation. Solving the appropriate equations of steady-state kinetics, we obtain the reaction velocity:

$$v = \frac{AS^2 + BS}{CS^2 + DS + K} \qquad (6.63)$$

where A, B, C, D, K are combinations of the kinetic constants. Expression (6.63) is similar to Eq. (6.49), despite the absence of cooperative interactions. The singularities on the $v(S)$ curve disappear if $k_3 = 0$ or $k_5 = 0$. Indeed, in some cases there are observed very large times of conformational changes of enzymes, of the order of minutes and even of tens of minutes. Such features on the $v(S)$ curves as an intermediate plateau are presumably accounted for by the nonequilibrium nature of enzyme systems. However, in most cases, the presence of inflections and even of maxima is thought to be due to the interaction of subunits in a protein that has a quaternary structure (see Sec. 6.8).

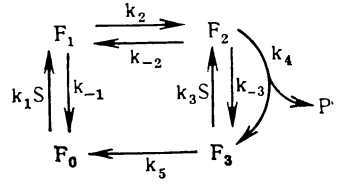


Fig. 6.20. Schematic representation of the Rabin reaction.

We have dealt above with only relatively simple enzymatic processes. In more complicated cases, calculations even under steady-state conditions become unwieldy—one has to solve a large number of simultaneous equations. For these problems there have been worked out algorithms, in particular those based on the *theory of graphs* (Goldshtein, Volkenshtein). In mathematics, a graph is a topological chart constructed from nodal points and lines between them. In essence, the diagrams shown in Figs. 6.17 and 6.20 are graphs. Algorithms of this kind allow one to obtain analytical expressions for reaction rates without solving kinetic equations.

It should be noted that in any more or less complicated cases the number of constants is so large that they cannot be found on the basis of experiments carried out under steady-state conditions. Nonsteady-state enzyme kinetics and, hence, the mechanism of enzyme action is effectively studied by the methods of *chemical relaxation* developed

by Eigen. The system is brought out of equilibrium or stationary state by rapidly changing an external parameter, and its approach to a new equilibrium or stationary state is studied. Most often use is made of a concentration or temperature jump, the action of ultrasound, etc. Relaxation times up to 10^{-10} s are accessible to measurement.

Let us consider the simplest one-step reaction (S is a ligand)

$$F_0 + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} F_1$$

The kinetic equation has the form

$$\dot{F}_0 = -k_1 S + k_{-1} F_1 \tag{6.64}$$

We have deviations from steady-state concentrations $\bar{F}_0$, $\bar{S}$, $\bar{F}_1$: $F_0 = \bar{F}_0 + x_0$, $F_1 = \bar{F}_1 + x_1$, $S = \bar{S} + y$. Equation (6.64) yields

$$\dot{x}_0 = -k\bar{F}_0 y - k_1 \bar{S} x_0 - k_1 x_0 y + k_{-1} x_1 \tag{6.65}$$

With a small perturbation we neglect $k_1 x_0 y$. From the laws of conservation of mass it follows that $x_0 = -x_1 = y$. Hence,

$$\dot{y} = -(k_1 \bar{F}_0 + k_1 \bar{S} + k_{-1})\, y \tag{6.66}$$

or

$$\dot{y} + \tau^{-1} y = 0 \tag{6.67}$$

where $\tau^{-1} = k_1 \bar{F}_0 + k_1 \bar{S} + k_{-1}$. The solution of Eq. (6.67) has the form

$$y = y\,(0)\, \exp\,(-t/\tau) \tag{6.68}$$

The value of y at time $t = 0$ is denoted as y (0). Measurement of τ at different concentrations $\bar{F}_0$ and $\bar{S}$ enables one to determine k_1 and k_{-1}.

We have made use here of the usual procedure of linearization of a nonlinear equation (see Sec. 15.2).

6.8. Myoglobin and Hemoglobin

Hemoglobin (Hb) and myoglobin (Mb) are not enzymes. Their function consists in the reversible binding of molecular oxygen. Myoglobin serves to store oxygen for its subsequent utilization. Therefore, large amounts of myoglobin are contained in the organism of whales existing predominantly under water. Hemoglobin is the functional protein of erythrocytes which serves for the transportation of oxygen from the lungs to all organs and tissues and which participates in the reverse transportation of CO_2.

The investigation of myoglobin and hemoglobin, however, yields information of great significance for the understanding of the prop-

erties of ordinary and allosteric electronic-conformational interactions. The binding of O_2 and other ligands by these proteins bears a close resemblance to the binding of the substrate to the enzyme. Molecular oxygen penetrates into the interior of myoglobin and hemoglobin molecules, but, in distinction to the substrate, it is not subjected to a chemical conversion. Sometimes, myoglobin and hemoglobin are called "honorary enzymes".

Myoglobin and hemoglobin are produced in crystalline form. Both proteins have been studied in detail by X-ray structure analysis

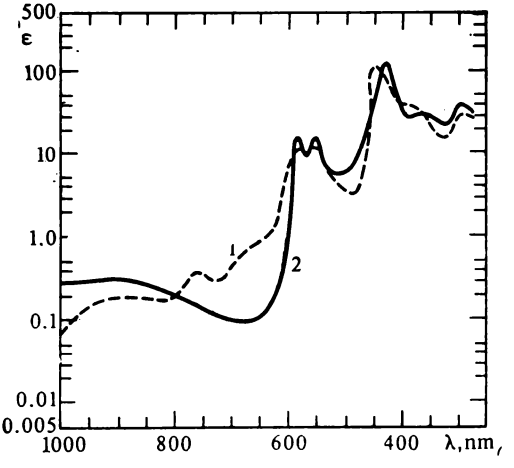


Fig. 6.21. The absorption spectra of Hb (*1*) and HbO_8 (*2*).

with a resolution of up to 0.28 nm, both in the oxygenated (MbO_2, HbO_8) and in the deoxygenated form (Mb, Hb). The presence in myoglobin and hemoglobin of the prosthetic groups of the *heme*, which has specific electronic properties, allows one to study these proteins by means of spectroscopy, electron paramagnetic resonance and also magnetic susceptibility, the Mössbauer effect (see page 152), etc.

Myoglobin has no quaternary structure (page 126); the hemoglobin molecule is composed of four subunits—two identical α-chains and two identical β-chains, each of which is similar to, but not identical with, the myoglobin molecule. Accordingly, hemoglobin, unlike myoglobin, displays cooperative properties—a so-called *heme-heme interaction* takes place in hemoglobin.

Oxygen and other ligands add on to the heme group, saturating the sixth coordination valence of the iron atom. The formation of this bond invokes a number of events in the protein molecule.

The heme is ferroprotoporphyrin (see Fig. 2.14). The iron atom, which is in the divalent ferro-state (Fe^{2+}), is coordinately bonded to the four nitrogen atoms of the pyrrole groups of the planar porphyrin ring. The fifth coordination bond directed perpendicularly to the ring plane joins the iron atom with histidine imidazole and the sixth valence is either free or occupied by a ligand.

Hemoglobin and myoglobin and their oxygenated forms contain ferroheme. The oxidation of hemoglobin and myoglobin leads to the formation of ferri-compounds that contain a trivalent iron atom, Fe^{3+}.

The iron atom is paramagnetic. Table 6.2 lists the characteristics of the magnetic properties of hemoglobin and its derivatives in various states.

Table 6.2. The Magnetic Properties of Hemoglobin

Compound	Valence	Sixth ligand	Magnetic moment, Bohr magnetons	Spin
Hemoglobin (Hb)	2	Absent	5.2-5.5	2
Oxyhemoglobin (HbO_8)	2	O_2	0	0
Carboxyhemoglobin [$Hb(CO)_4$]	2	CO	0	0
Ferrihemoglobin (methemoglobin, metHb)	3	H_2O	5.6-5.8	5/2
Hydroxymethemoglobin	3	OH^-	4.5-4.7	1/2, 5/2
Ferrihemoglobin azide	3	N_3^-	2.4-2.8	1/2
Ferrihemoglobin cyanide	3	CN^-	2.3-2.5	1/2

We see that ferroheme can be in a high-spin (Hb) and a low-spin [HbO_8, $Hb(CO)_4$] state. Ligand binding manifests itself in the absorption spectrum in the visible region (Fig. 6.21). Table 6.3 gives the corresponding data.

Table 6.3. The Absorption Spectra of Some Selected Heme Compounds

Compound	α-Band		β-Band		Soret band	
	λ, nm	$\varepsilon \cdot 10^{-3}$	λ, nm	$\varepsilon \cdot 10^{-3}$	λ, nm	$\varepsilon \cdot 10^{-3}$
Heme	565	6.1			390	39.6
CO-heme	562	14.6	530	11.9	406.5	147
Hb	555	13.5			430	119
HbO_8	577	14.6	542	13.8	412	135
$Hb(CO)_4$	569	13.4	539	13.4	419	191
Mb	555	12.0			435	114
MbO_2	582	13.1	544	12.7	417	119
MbCO	578	12.3	541	14.1	423	185

The long-wave bands of heme, Hb and Mb, are not usually called α-bands.

The interpretation of the magnetic and spectral properties of heme is based on quantum-mechanical analysis. The outer electrons of the iron atom have a $3d^6$ configuration for Fe^{2+} and a $3d^5$ configuration for Fe^{3+} (Fig. 6.22). Theoretical calculations are based on the ligand field theory. The molecular orbitals of the system are represented by the linear combinations of the orbitals of Fe, porphyrin ring, and ligands. Calculations give an electron-density distribution and electron energy levels. A number of questions, however, have not yet been solved, in particular, the problem of the strained heme with five coordination bonds in Hb and Mb.

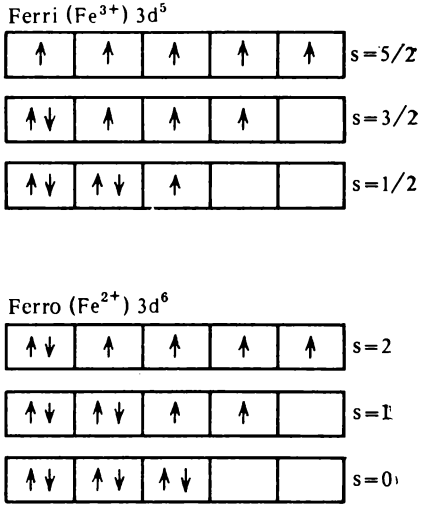


Fig. 6.22. The electronic structure of Fe^{2+} and Fe^{3+}.

Hemoglobin and myoglobin are paramagnetic, in contrast to the vast majority of other biological molecules. In this connection, mention should be made of the so-called *magnetobiology*, which is concerned with study of the effect of magnetic fields on biological phenomena. Magnetic fields could, in principle, have an effect on the behaviour of hemoglobin and myoglobin and of the corresponding cells and also on the kinetics of biochemical reactions that involve free radicals. However, no reliable data have yet been obtained on the effect of a magnetic field on biological phenomena.

A special place is occupied by bacteria that have recently been discovered in peat bogs in the United States and that contain ferrimagnetic magnetite, $FeO \cdot Fe_2O_3$. These bacteria are capable of orientation in a magnetic field and they can be magnetized. At the same time, there is no scientific ground for stating that "magnetized water" plays a biological role, that water that has passed through a magnetic field speeds up, as it were, the growth of plants, has bactericidal properties, etc.

We know the primary structures of the hemoglobin and myoglobin of many species of animals and also of a large number of human mutant hemoglobins (see page 36). The deciphering of the spatial structure of Hb and Mb and the elucidation of conformational changes that arise upon binding of ligands are of particular significance. It is exactly for myoglobin and hemoglobin that the relation between

structure and properties has been especially thoroughly studied.

The curve of saturation of Mb with molecular oxygen, $\overline{Y}(p)$, where p is the partial pressure of O_2, is similar to the Langmuir isotherm:

$$\overline{Y} = \frac{p}{K+p} \tag{6.69}$$

On the contrary, the plot of $\overline{Y}$ versus p for hemoglobin has a point of inflection—it is a sigmoid curve. It can be described by the *Hill equation*:

$$\overline{Y} = \frac{p^n}{K'+p^n} \tag{6.70}$$

where K' is a constant and the parameter $n = 2.8$. The two curves are shown in Fig. 6.23 [$1 - \overline{Y} = Kp/(1 + Kp)$; $2 - \overline{Y} = Kp^n/(1 + Kp^n)$; $n = 2.8$].

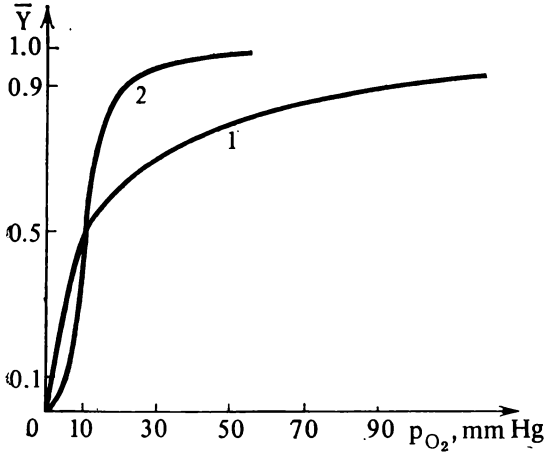


Fig. 6.23. Saturation curves for myoglobin (*1*) and hemoglobin (*2*) saturated with molecular oxygen.

The physiological meaning of the sigmoid curve of $\overline{Y}$ vs. p for hemoglobin is that the affinity of Hb for O_2 decreases as molecular oxygen is eliminated. The partial pressure of O_2 in tissues does not undergo significant changes. If the hyperbolic curve of $\overline{Y}$ vs. p [Eq. (6.69)] had been characteristic of hemoglobin, only a small fraction of molecular oxygen transported would have been released in tissues. As a result, the organism would have suffocated even in the atmosphere of pure oxygen. The efficiency of respiratory transport

is also controlled by the presence of a cofactor—2,3-diphosphoglycerate (DPG), which reduces the affinity of hemoglobin for O_2, and by the Bohr effect.

The *Bohr effect*, which is characteristic of hemoglobin but not of myoglobin, consists in the dependence of the affinity of Hb for O_2 on the pH of the medium. The affinity is minimal near pH 6 and maximal near pH 9. In other words, small concentrations of proton facilitate the addition of O_2 and, conversely, small concentrations of O_2 facilitate the addition of protons to Hb. By increasing the pH of venous blood and, thereby, enhancing its ability to absorb bicarbonate, the Bohr effect provides the principal mechanism of reverse transport of CO_2 from tissues to lungs.

The difference of the coefficient n from unity in the Hill equation (6.70) and the corresponding sigmoidicity of the curve of $\bar{Y}$ vs. p reflects the heme-heme interaction, i.e., the interplay of the four subunits and, hence, the cooperativity of the addition of O_2. The changes in enthalpy upon binding of the first, second, third, and fourth O_2 molecules by sheep hemoglobin are as follows: $\Delta H_1 = = -65.9 \pm 3.3$, $\Delta H_2 = -47.8 \pm 10.5$, $\Delta H_3 = -32.7 \pm 13.8$, and $\Delta H_4 = -36.5 \pm 13.8$ kJ/mole. The binding of oxygen O_2 is increased as the addition is continued.

The events that occur in the hemoglobin molecule on oxygenation have been disclosed by Perutz by means of X-ray diffraction studies. The O_2 molecule binds to the Fe atom of heme. In HbO_8 the Fe atom is in the heme plane, at its centre. In high-spin hemoglobin, the Fe atom is at a distance of 0.05 nm from this plane in the direction of the imidazole ring of His *F*8. In such a state the coordination number of Fe is 5. Oxygenation makes Fe pass to a low-spin state and increases the number of Fe ligands by unity. These changes cause alterations in contacts between the heme and the tightly packed amino-acid residues of the protein. In other words, electronic-conformational interactions result in a rearrangement of the protein globules.

The heme is in direct contact with 60 atoms of the protein. Introduction of even the smallest ligand, OH^-, which has a radius of 0.15 nm, brings about a conformational change in the β-subunits of hemoglobin. In these subunits the γ-CH_3 group of the residue Val *E*11 is found to be a distance of 0.25 nm from OH^-, which is less than the sum of the van der Waals radii. Hence, in the β-globules of Hb there is no place even for the smallest ligand, and on oxygenation the distance between the heme and Val *E*11 must increase by about 0.1 nm. On the contrary, in the α-subunits no such displacement occurs, since the width of the "pocket" is enough for the ligand to be incorporated.

In HbO_8 the C-terminus ends of all the four chains are completely free to rotate, and the second-to-last residues Tyr (140) are partially free. On the contrary, in Hb each of the C-ends is doubly fixed by

salt bridges of Arg (141) α_1 with Asp (126) α_2, His (146) β_1 with Lys (40) α_2 and Asp (84) β_1. All the four penultimate Tyr residues are rigidly fixed in the cavities between the *F*- and *H*-helices by van der Waals and hydrogen bonds. The oxygenation of Hb causes a shift of the *F*-helix and a rupture of salt bridges, as a result of which Tyr (140) is removed from the pocket between the *F*- and *H*-helices; at the same time, the width of the pocket is diminished by 0.13 nm in the α-chains and by 0.2 nm in the β-chains. This rearrangement is

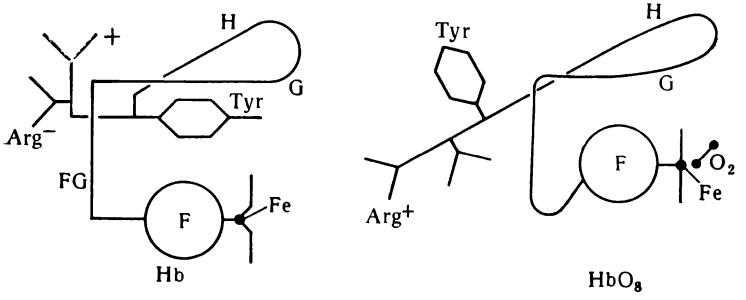


Fig. 6.24. The displacement of Tyr HG2 (140) on oxygenation.

shown in Fig. 6.24. The rearrangements in the subglobules and ruptures of the salt bridges alter the quaternary structure. The displacements that occur in the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ contacts are 0.1 nm and those in the $\alpha_1\beta_2$ and $\alpha_2\beta_1$ contacts are 0.7 nm, which is a very large distance. These latter contacts are of greatest significance. The mutant replacements in the corresponding regions appreciably reduce the heme-heme interaction.

The term "heme-heme interaction" is, of course, tentative. What is meant here is not the interaction between heme groups—the distance between them is too large for such an interaction to be realized. The interaction of the subunits is determined by conformational events, which are triggered by the displacement of the Fe atom into the plane of the porphyrin ring and by the corresponding shift of the proximal His by 0.075-0.095 nm. The *F*-helix is shifted to the centre of the molecule and pushes Tyr (140) from the cavity between the *F*- and *H*-helices. The Tyr residue that has been expelled drags Arg (141) and disrupts the salt bridges with the opposite α-chain. The situation in the β-chains is different. Before the ligand reaches the Fe atom, it has to open up the cavity near the heme. There is formed a new Fe—ligand bond, the Fe atom is moved into the plane of the heme, and the *F*-helix shifts to the centre of the molecule and

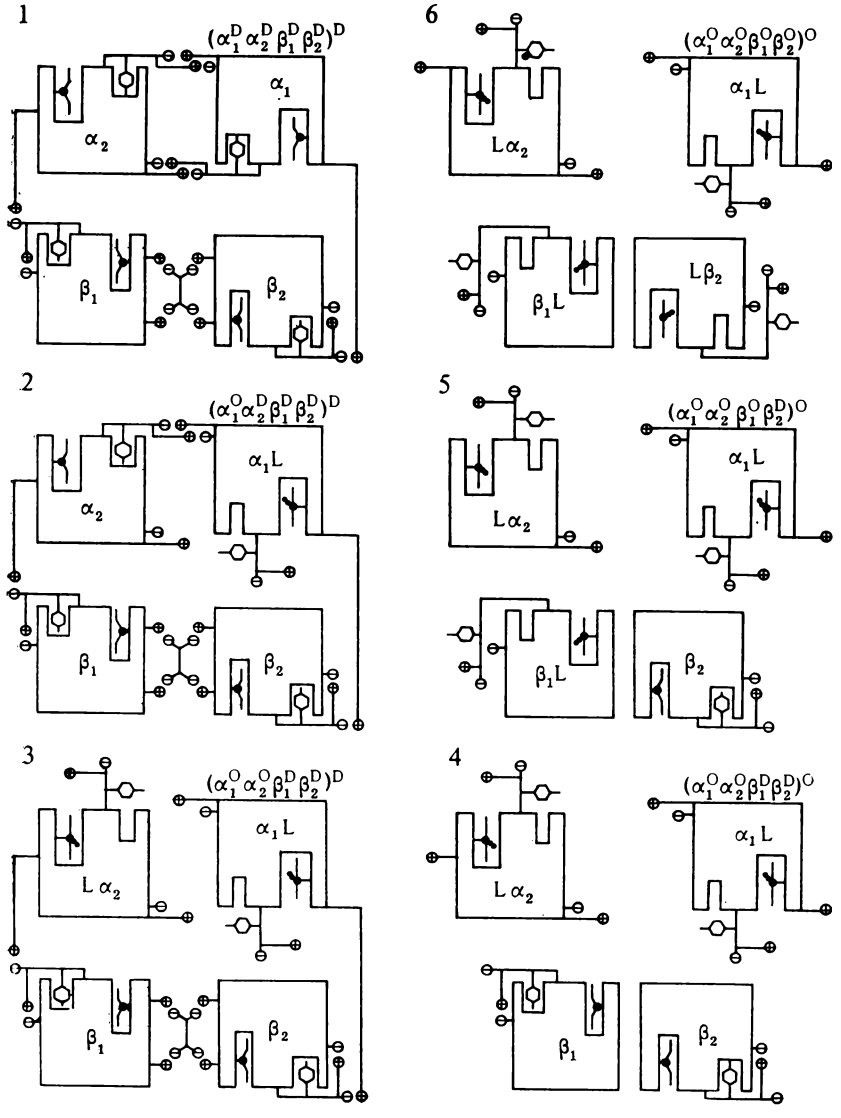


Fig. 6.25. The oxygenation of hemoglobin according to Perutz:
1. Hemoglobin with intact salt bridges and with a molecule of diphosphoglycerate sandwiched between two β-chains. *2.* HbO_2. *3.* HbO_4. The steps 1-2 and 2-3 involve the oxygenation of the α-chain. *4.* HbO_4 with an altered conformation; the steps 3-4 involve a conformational change. *5.* HbO_6. *6.* HbO_8.

pushes Tyr (145) from its pocket. This residue pulls His (146) and disrupts its salt bridge with Asp (94).

The structure of Hb is stabilized by DPG, which forms extra salt bridges between the β-subglobules. On oxygenation DPG is removed from the molecule.

According to Perutz, each of the subunits can be in the state of *deoxy-* and *oxy*-conformation. The binding of O_2 makes the subunit pass to the oxy-conformation, but the quaternary structure remains to be deoxy until two O_2 molecules are bound; oxygenation presumably starts from α-subglobules, since there is sufficient place in them for a ligand to be incorporated. The scheme of the process shown in Fig. 6.25 is

$$(\alpha_1^D\alpha_2^D\beta_1^D\beta_2^D)^D \xrightarrow{+O_2} (\alpha_1^O\alpha_2^D\beta_1^D\beta_2^D)^D \xrightarrow{+O_2} (\alpha_1^O\alpha_2^O\beta_1^D\beta_2^D)^D \xrightarrow{-DPG}$$

$$\rightarrow (\alpha_1^O\alpha_2^O\beta_1^D\beta_2^D)^O \xrightarrow{+O_2} (\alpha_1^O\alpha_2^O\beta_1^O\beta_2^D)^O \xrightarrow{+O_2} (\alpha_1^O\alpha_2^O\beta_1^O\beta_2^O)^O$$

The essence of the process is as follows: as a result of its specific structure, the heme group enhances a slight change in the atomic radius that occurs in the Fe atom upon transition from the high-spin to the low-spin state and transforms this change to a large displacement of the heme-bound His. The electronic events are transformed to conformational events, with all the consequences involved.

The interaction energy of the subunits is about 50 kJ/mole, this corresponding to the energy of the salt bridges that form the contacts.

Hemoglobin is tetrameric, its splitting into dimers occurs at small concentrations. This must entail the breaking of the salt bridges and the quaternary structure will assume the oxy-conformation. Cooperative effects are determined by the tetrameric structure.

Valuable information about interactions in Hb has been obtained by means of magnetic polarimetry (Sec. 5.8) and the Mössbauer effect (Sec. 5.3). Magnetic rotatory dispersion (MRD) and magnetic circular dichroism (MCD) are extremely sensitive to the specific features of the structure of Hb and Mb which are practically indistinguishable in the absorption spectra. Figure 6.26 shows MRD curves for Mb and its complexes with ligands. These are electronic effects. Their good correlation with the conformational stability of the complexes on denaturation with urea is a direct evidence of electronic-conformational interactions (Sharonov *et al.*).

The α-band of Mb, which is not practically undetectable on absorption but observable in MRD and MCD, is very sensitive to the heme-heme interaction. For Mb and also for the α- and β-subunits of Hb in the α- and β-bands the MRD curve is about the same. On the contrary, in tetrameric Hb the effect in the α-band is twice as strong as in the β-band (Fig. 6.27). The MRD method has made it possible

to study the dissociation of Hb into subunits. It has been found that at pH 10-11 the tetramers dissociate into noncooperative $\alpha\beta$ dimers.

The Bohr effect has been interpreted in molecular terms. The conformational transitions oxy-Hb $\rightleftarrows$ deoxy-Hb alter the environment of three pairs of weak bases in such a manner that they tend to add

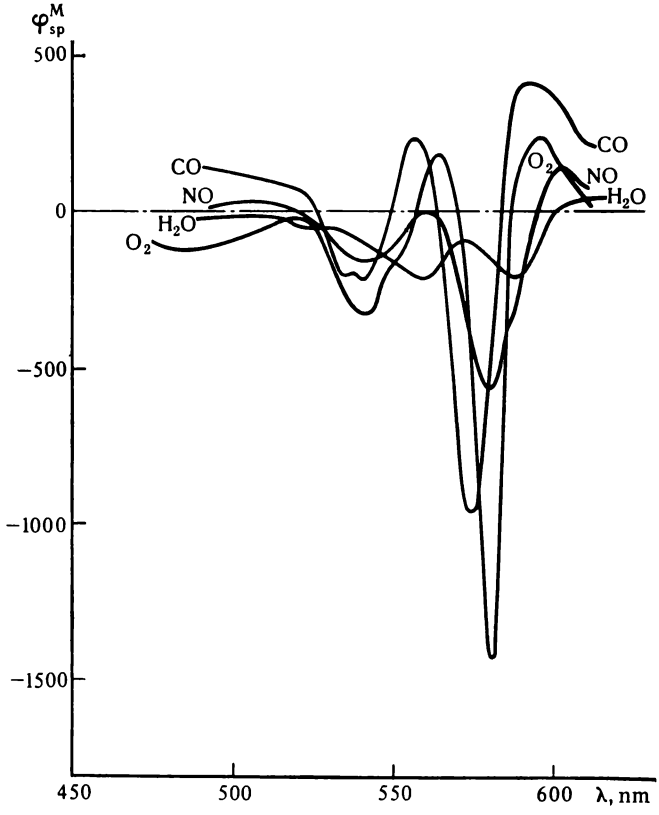


Fig. 6.26. The magnetic-rotatory-dispersion (MRD) curves of Mb and its complexes.

protons upon expulsion of O_2. This is accounted for by transfers of ionogenic amino-acid residues. The ratio of O_2 to protons in Hb is the same as the substrate-allosteric effector ratio in allosteric enzymes.

Wyman has worked out a phenomenological theory of bound functions for hemoglobin and allosteric enzymes. Let us assume that the molecule of an allosteric enzyme binds two ligands, X and Y,

while having q active sites for X and r active sites for Y. The total concentration of molecules of the allosteric enzyme is given by

$$c = c_0 \sum_{i=0}^{q} \sum_{j=0}^{r} K_{ij} x^i y^j \tag{6.71}$$

where c_0 is the concentration of free molecules, x and y are the con-

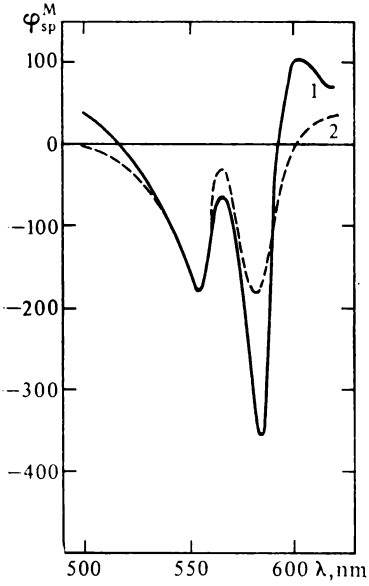


Fig. 6.27. The MRD curves of Hb (*1*) and isolated α- and β-chains (*2*).

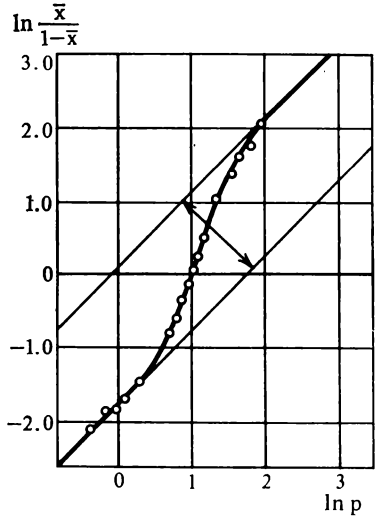


Fig. 6.28. The Hill graph for the saturation of the horse hemoglobin with oxygen.

centrations of X and Y, respectively, and K_{ij} is the equilibrium constant for the reaction

$$M + iX + jY \rightleftarrows MX_iY_j$$

The saturation functions are given by

$$Y_X = \frac{\partial \ln \sum\sum K_{ij} x^i y^j}{\partial \ln x} \quad \text{and} \quad Y_Y = \frac{\partial \ln \sum\sum K_{ij} x^i y^j}{\partial \ln y} \tag{6.72}$$

From this it follows that

$$\left(\frac{\partial \ln Y_X}{\partial \ln y}\right)_x = \left(\frac{\partial \ln Y_Y}{\partial \ln x}\right)_y \tag{6.73}$$

It is convenient to represent the equilibrium curve of the ligand not by the quantity Y_X but in the form of the plot of $\ln [\bar{x}/(1 - \bar{x})]$, where $\bar{x} = Y_X/q$, against $\ln x$ and to introduce the parameter n defined as

$$n = \frac{d \ln [\bar{x}/(1-\bar{x})]}{d \ln x} = \frac{1}{\bar{x}(1-\bar{x})} \frac{d\bar{x}}{d \ln x} \tag{6.74}$$

If all the sites are identical and are independent of each other, then the plot is a straight line with $n = 1$. For stabilizing interactions, i.e., at $(\partial Y_X/\partial \ln x) > q$, the value of n is greater than unity. This is in accord with the Hill equation (6.70), which can be rewritten in the following form:

$$\ln \frac{\bar{x}}{1-\bar{x}} = \ln K + n \ln x \tag{6.75}$$

Figure 6.28 shows the graph described by Eq. (6.75), where $x = p$, for horse hemoglobin. Here $n = 2.95 \pm 0.05$ and the free energy of interaction of Hb with O_2 is 10.9 kJ/mole. Using the Wyman theory, one can easily describe the Bohr effect ($X = O_2$ and $Y = H^+$).

Hemoglobin can bind O_2 and CO. Upon saturation the distribution coefficient A is independent of the partial pressures of the two gases (*the first Haldane law*):

$$\frac{[Hb(CO)_4]}{[HbO_8]} = A \frac{p_{CO}}{p_{O_2}} \tag{6.76}$$

The second Haldane law states that the saturation upon interaction with a mixture of O_2 and CO is a function of $p_{O_2} + Ap_{CO}$. This law follows from Eq. (6.76).

The properties of hemoglobin demonstrate the dynamic behaviour of the protein. The tertiary and quaternary structures of hemoglobin oscillate rapidly and continuously between the oxy- and deoxy-conformations. The addition of a ligand causes a shift of the conformational equilibrium. This is what we know as the electronic-conformational interaction.

Chapter 7 Physics of Nucleic Acids

7.1. Molecular Biology and Physics

Molecular biology is concerned with study of the molecular nature of basic life phenomena, primarily heredity and variability. These phenomena are determined by the structure and properties of nucleic acids—informational macromolecules. The origin and development of molecular biology is linked with the discovery of the genetic role of nucleic acids and with its deciphering. The *genes*, i.e., fragments of DNA and RNA molecules, control the synthesis of proteins. These molecules perform "legislative" functions, while proteins are "executive" compounds. Molecular biology was originated with the discovery of the *transformation of bacteria* by means of DNA (Avery, MacLeod, McCarty, 1944). Molecular biology seeks an explanation for the biological phenomena in chemistry and molecular physics. It deals with a wide range of life processes, including enzyme catalysis, membrane transport, mechanochemical phenomena, etc. In distinction to classical biochemistry, molecular biology combines with physics and its specificity consists exactly in the physical aspects of investigations and problems.

In the preceding chapters we examined the problems of molecular biophysics connected with the structure and functionality of proteins. We now turn to physical problems pertaining to the structure and properties of nucleic acids, to protein biosynthesis. These problems are as follows.

1. The secondary and tertiary structures of all kinds of nucleic acids. The relation between structure and biological properties.

2. The genetic code which determines the relationship between the nucleotide sequence in genetic DNA and RNA and in messenger RNA (mRNA) and the amino-acid sequence in the protein chain. The problem of the genetic code, which arises in the atomic-molecular treatment of the processes of protein biosynthesis, is physical in nature. This problem has been solved.

3. The physical meaning and origin of the *genetic code*. The code is not accidental; presumably, it originated in the course of prebiological evolution. Physics is called upon to disclose the internal

relations of the code and to build a theoretical model of its evolution.

4. The conformational properties of nucleic acids. The behaviour of nucleic acids, like that of proteins, is monitored by electronic-conformational interactions. Theoretical and experimental investigations of the conformations of nucleic acids and their changes unravel the mechanisms of molecular biological processes. This refers, in particular, to denaturation and renaturation of nucleic acids, DNA reduplication and replication (*transcription* and *reverse transcription*) of DNA and RNA.

5. Internucleic and nucleic-protein molecular recognition. The nucleic-acid-protein interaction is of greatest significance for the regulation of protein biosynthesis and of the action of the genes.

6. The interactions of nucleic acids with small molecules and ions which are very important for the functioning of nucleic acids and for the formation of higher levels of their structure.

7. The structure and properties of supermolecular nucleic-acid-protein system—*ribosomes* and *chromosomes*. The molecular mechanisms of their functioning.

8. The physical mechanisms of point mutations, i.e., alterations in the primary structure of informational macromolecules of DNA and RNA.

9. The theoretical (physico-mathematical) modelling of development processes—prebiological and biological evolution, *ontogenesis* and *cancerogenesis*—which lean on the known properties of nucleic acids.

This list does not claim to be complete. Some of the problems mentioned above have already been solved; others are still in the initial stage of resolution.

Physics has played an important role in the building of molecular biology. It will suffice to point to the discovery of the double helix of DNA made with the aid of X-ray diffraction analysis, and the clear-cut formulation of the problem of the genetic code. The general physical approach to biological problems used by Schrödinger in his book *What is Life*? (1945) stimulated the development of molecular biology.

For the last 10-15 years the situation in molecular biology has changed. After the biological molecules had been studied and the genetic code deciphered molecular biology switched over to much more complex supermolecular and cellular systems. It has proved to be possible to approach to the solution of problems associated with the molecular genetics of *eukaryotes*, with the phenomena of ontogenesis. At this stage molecular biology has divorced from theoretical and experimental molecular physics. This happened because of the complexity of the processes studied by modern molecular biology. Only in rare cases is the information on these processes sufficient for the formulation of physical problems. However, the physico-

mathematical modelling of the behaviour of eukaryotes (*mitosis* and *meiosis*) and ontogenesis is being developed and serves as an important source of ideas for theoretical biology. We shall introduce the reader to these trends of investigations in the last chapters of the book.

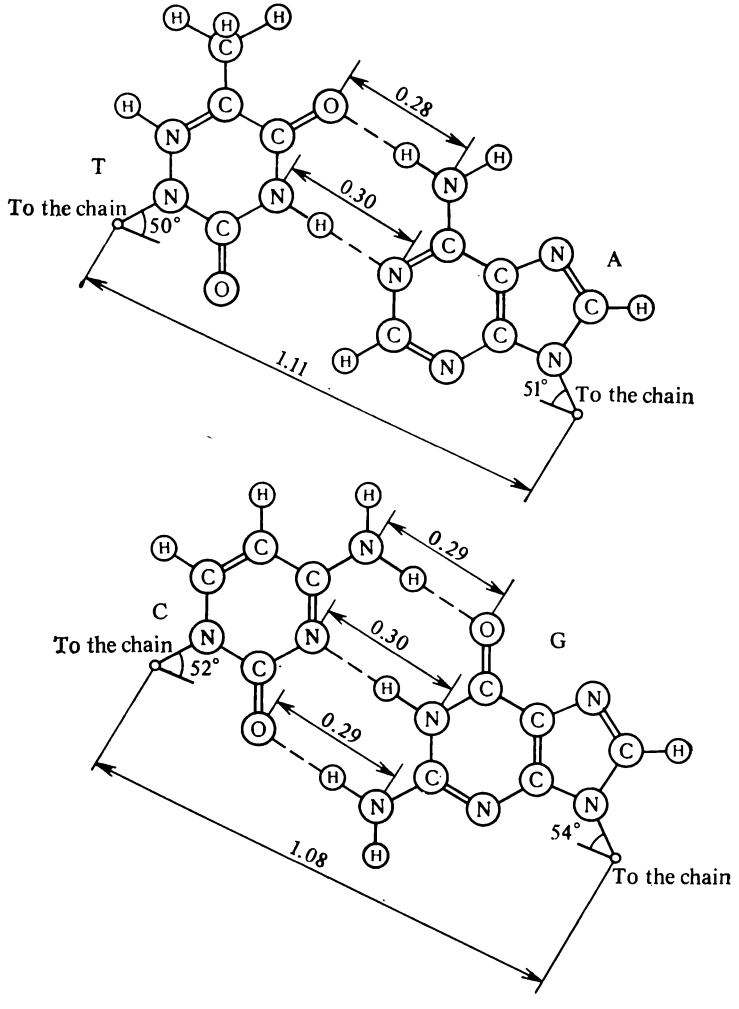


Fig. 7.1. Adenine-thymine (A-T) and guanine-cytosine (G-C) pairs.

7.2. The Structure of Nucleic Acids

The relation between the structure of the molecule and its functional properties is far from being always simple and obvious. We have seen how difficult it is to establish this relation for proteins. The situation is different with nucleic acids, in particular, with DNA. Here at least one highly important function—DNA replication—has been qualitatively explained immediately after the discovery of the secondary structure.

The secondary structure of DNA was established by X-ray diffraction analysis in the works of Franklin, Crick, Watson, and Wilkins (1952). The oriented fibres of lithium salts of native DNA gave X-ray diffraction patterns containing up to 100 reflections (see Fig. 5.5). The cross-shaped array of reflections directly shows that the structure is helical.

Native DNA is built up in the form of a *double helix*, which is composed of two intertwined polynucleotide chains, whose nitrogenous bases are bound in pairs by hydrogen bonds, adenine (A) always pairing with thymine (T) and guanine (G) always pairing with cytosine (C). These pairs (*Watson-Crick pairs*) are shown schematically in Fig. 7.1 and the molecular model of the double helix (the *B*-form) is given in Fig. 7.2. Thus, the two chains of DNA in the double helix are complementary to each other, i.e., there exists a unique correspondence between their nucleotides: T is complementary to A and G is complementary to C. This complementarity discloses the meaning of Chargaff's rules (see page 41).

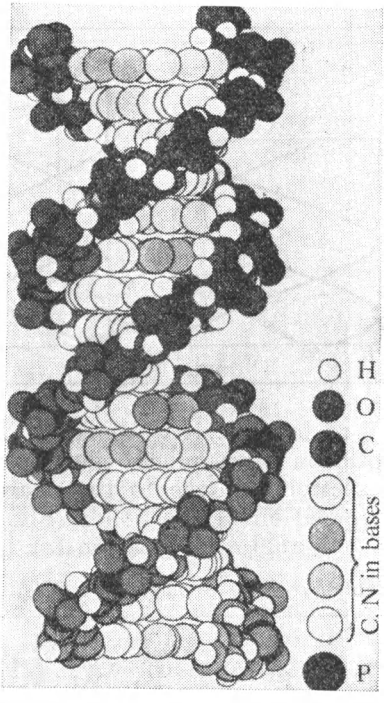


Fig. 7.2. The molecular model of the *B*-form of DNA.

DNA can crystallize in various double-stranded forms. At a low relative humidity (up to 70 %) DNA crystallizes in a monoclinic *A*-form with the lattice parameters being as follows: $a = 4.04$ nm, $b = 2.207$ nm, and $\beta = 97.1°$. At a high humidity a hexagonal *B*-form is obtained, for which $a = 4.6$ nm. The two forms of the

Na-salt of DNA are shown schematically in Fig. 7.3. The *C*-form was discovered later. Table 7.1 lists the characteristics of the *A*-, *B*-, and *C*-forms of DNA. They differ in pitch of the double helix, the angles of rotation between neighbouring base pairs, and the slope of the plane of base pairs to the helix axis.

The conformation of a polynucleotide chain is determined by a set of angles of rotation shown in Fig. 7.4. The angles are counted

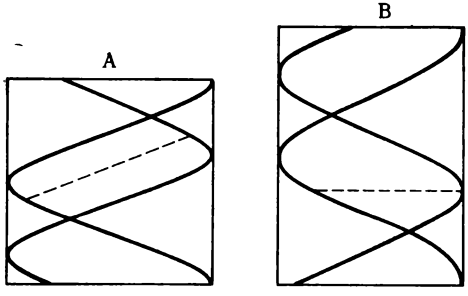


Fig. 7.3. The structure of the *A*- and *B*-forms of the Na-salt of DNA.

off as follows: if we look in the direction of the bond about which rotation takes place, the farthest bond is rotated in the clockwise direction relative to the nearest one. Counting is made from the *cis*-position of the bonds (Fig. 7.5.). Table 7.2 presents the values of the angles of rotation for the three conformations of DNA.

Table 7.1. The Geometry of DNA in Its Various Forms

DNA salt	Humidity, per cent	Number of units per turn of helix	Helix pitch, nm	Shift by one nucleotide, nm	Rotation by one nucleotide, deg	φ_1, deg	φ_2, deg
Na, *A*-form	75	11	2.82	0.255	32.7	20	16
Na, *B*-form	92	10	3.46	0.346	36	—	—
Li, *B*-form	66	10	3.37	0.337	36	2	5
Li, *C*-form	66	9.3	3.10	0.332	39	6	10

Here φ_1 is the angle between the normal to the helix axis and the plane of the bases; φ_2 is the dihedral angle between the planes of the paired bases.

The conformation of the polynucleotide chain depends on the conformation of the sugar—ribose or deoxyribose. The sugar is present in four conformations, which differ in which of the atoms of the five-membered ring, C_2 or C_3 (Fig. 7.4), is brought out of the plane of the other four atoms. If the atom C_2 (or C_3) is on the

same side of the plane as C_5, the conformation is called the *endo* conformation; in the opposite case we have the *exo* conformation. Hence, there are possible the C_2-*endo*, C_3-*endo*-, C_2-*exo*-, and C_3-*exo*-forms. The first two forms are encountered more frequently.

Theoretical calculations of the conformations of polynucleotide chains and double helices are not simple, since these conformations depend on many angles. The ribose or deoxyribose group is also

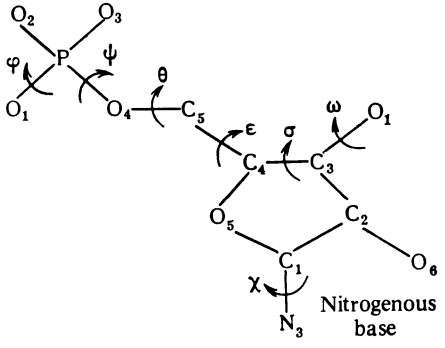


Fig. 7.4. The angles of rotation in a unit of the polynucleotide chain.

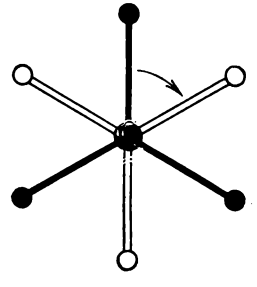


Fig. 7.5. The direction of rotation.

conformationally labile. Accordingly, calculations are carried out with a "rigid" and a "soft" sugar.

Experiment (the diffuse scattering of X-rays, CD, etc.) and theory have shown that one should speak not of the individual conformations *A*, *B*, *C* of DNA but rather of the conformational families

Table 7.2. The Conformations of DNA

Form	ψ, deg	θ, deg	ε, deg	σ, deg	ω, deg	φ, deg
A	283	167	67	76	221	279
B	281	212	58	130	147	282
C	315	143	48	168	211	212

A and *B*. The *C*-form as well as the *T*- and *D*-forms are similar to the *B*-form and belong to the *B*-family. The double helix proves to be a structure capable of continuously undergoing conformational changes. Within the *A* and *B* families the conformations are changed non-cooperatively, perhaps through a smooth gradual change in parameters of the helix. On the contrary, the change from one family

to the other, $A \rightarrow B$, occurs sharply, cooperatively, and is accompanied by a jumpwise change of the conformation of the sugar from 3-*endo* to 2-*endo*. It has been shown that for a given arrangement of the nitrogenous bases there is a limited number of conformations of the sugar-phosphate backbone of the regular double helix (Ivanov and coworkers).

All the above-mentioned double helices of nucleic acids are *right-handed.* However, *left-handed double helices* have recently been dis-

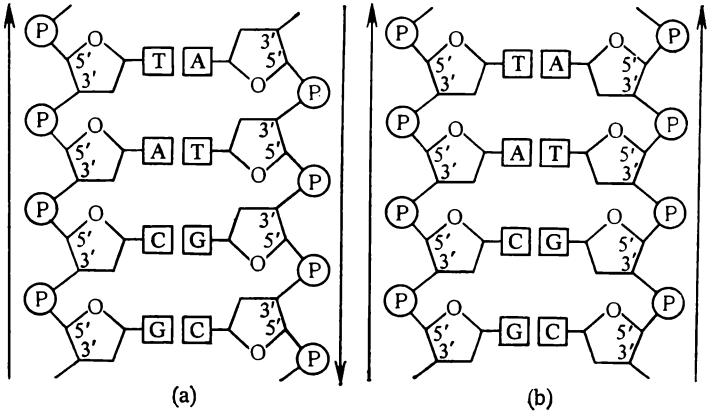


Fig. 7.6. Double helices built up of antiparallel (*a*) and parallel (*b*) chains.

covered. Such is the structure of the synthetic double-helical polymer polydeoxy (GC) and also of the double helix of poly-deoxy (AC)-poly-deoxy(GT).

The double helix of DNA in any possible conformation is characterized by wide and narrow helical grooves on the surface of the macromolecule.

Obviously, there are two possibilities for the organization of the double helix—from parallel and antiparallel chains (Fig. 7.6). The validity of the second model has been proved by X-ray diffraction and biochemical data. The chains in the double helix are antiparallel.

The specific double-stranded structure of DNA provides a direct explanation for two important facts—the replication of DNA upon mitosis and the metabolic stability of DNA. According to Crick and Watson, on replication the double helix is first split as a result of the breaking of the hydrogen bonds and uncoiling of the chains. Each strand acts as a template for the formation of a new strand, which is complementary to the template. The monomers of the new

chain join with the parental strand, forming Watson-Crick pairs A—T, T—A, G—C, and C—G. At the same time the nucleoside triphosphates are polymerized, this resulting in the formation of two double helices which are identical to the parental helix. Such a *semi-conservative model* (the new helix contains one old and one new chain) is indeed confirmed by experiment (see Sec. 7.7). The errors on replication are *point mutations* (Sec. 8.7). The double helix is similar to a one-dimensional crystal; the nitrogenous bases in it are tightly

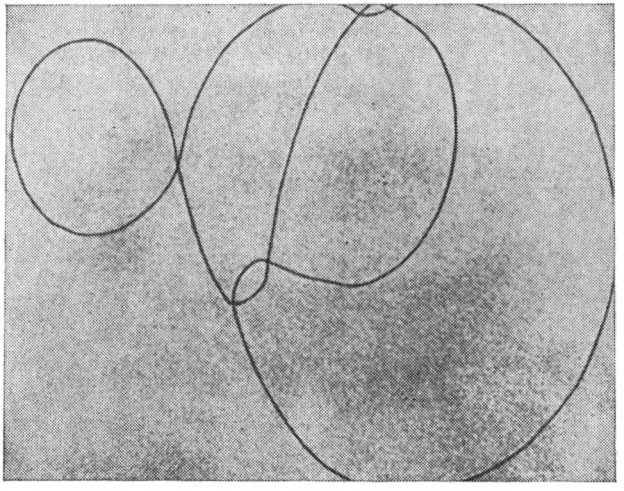


Fig. 7.7. Electron micrograph of the closed double helix of DNA from phage λ. Magnification 44,000×.

packed and strongly bound by weak interactions (Sec. 7.3). This accounts for the metabolic stability of DNA and the high conservation of the genes. In his time Schrödinger speculated on the factors responsible for the stability of a substance built up of the light atoms C, H, N, O, P. The discovery of the double helix provided an answer to this question.

DNA are the largest molecules known to science. From the T2 phage there were isolated circular double-stranded molecules of DNA with a length of 49 μm, and the DNA molecules produced from *E. coli* were 400 μm long, which corresponds to a molecular weight of the order of 10^9. Figures 7.7 and 7.8 show the photographs of the DNA molecules obtained in the electron microscope.

Solutions of native DNA are highly viscous. This does not mean, however, that the entire macromolecule is a double-helical rigid

rod. DNA is twisted into a very loose coil, since its rigidity is very great—the *persistent length* of DNA (see page 83) is 50 nm in 0.15*M* NaCl and 80 nm in 0.0014*M* NaCl.

The length 50 nm corresponds approximately to 150 nucleotide units. The density of DNA, which is approximately equal to 1.7 g/cm^3,

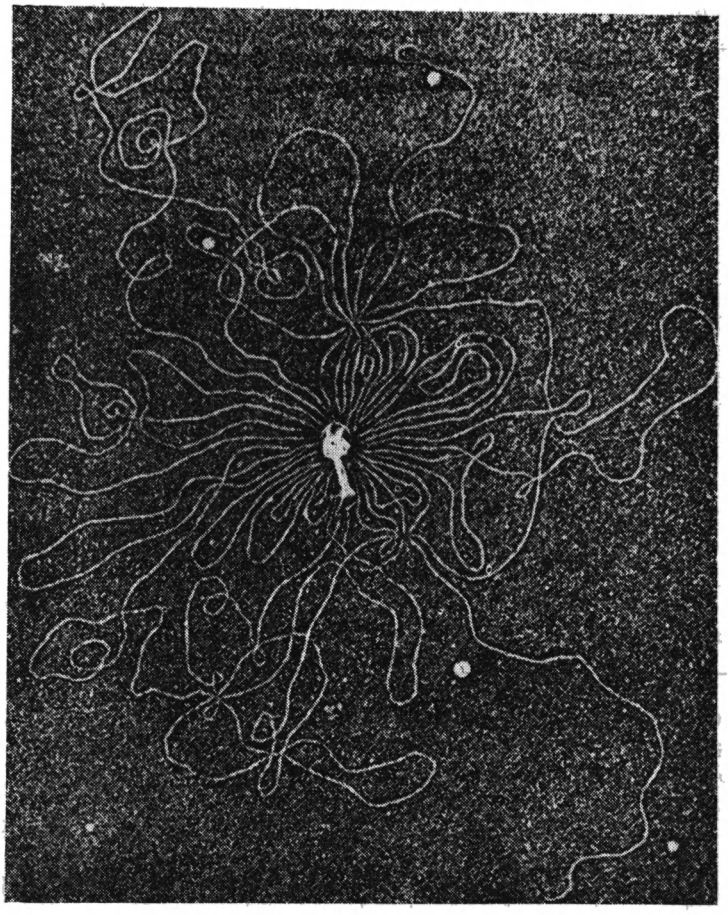


Fig. 7.8. Electron micrograph of DNA from phage T2. Magnification 80,000×.

increases linearly with increasing content of G-C pairs (the molecular weight of A-T is 247 and that of G-C is 273). However, the A-T and G-C pairs occupy nearly the same volumes in the double

helix. For each percent of the G-C pair there is an increase in density of 0.00103 g/cm^3.

On heating, change of pH, etc., DNA undergoes denaturation—the transition from the double helix to two coils (Sections 7.4 and 7.5). This transition can conveniently be followed from the absorption spectrum of DNA (in the 260 nm region) and from CD curves. The double helix exhibits high hypochromism in this region—the absorption intensity of a helical nucleic acid is much smaller than that of the denatured, coiled structure.

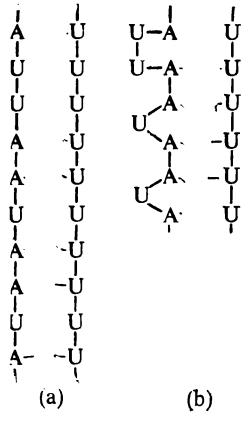


Fig. 7.9. The possible secondary structures of poly (U)-poly(AU) complexes.

The circular dichroism (CD) of oriented and unoriented films of DNA is changed with a change in humidity and ionic strength. The CD spectra of the Na- and Li-salts of DNA with a relative humidity equal to or greater than 92 percent are the same as those of solutions and correspond to the *B*-form. With a relative humidity ranging from 75 to 66 percent the films with 3 percent of Na^+ give a CD spectrum that corresponds to the *A*-form. Circular dichroism yields information on the conformational properties of DNA in solutions and in films. Very interesting are the investigations of the anisotropy of CD and optical activity of DNA macromolecules oriented in the stream (page 173).

The above-indicated methods disclose two-stranded portions in RNA molecules, in which no complementary chains are present and which do not obey Chargaff's rules. The double-stranded portions of nucleic acids are modelled by synthetic polynucleotides. In 0.1*M* NaCl, poly-A forms a double helix with poly-U, the strongest hypochromic effect (i.e., the greatest degree of double-helicity) being observed with a 1 : 1 composition of the polynucleotide mixture. In the presence of divalent cations, in 1.2×10^{-2} *M* solutions of $MgCl_2$, maximum hypochromicity corresponds to the 2 : 1 ratio of poly-U and poly-A, i.e., there is formed a triple helix. Cations are necessary for the formation of a helix, since they compensate for the charges on the phosphate groups and, hence, diminish the electrostatic repulsion of the chains.

A study of mixtures of poly-U and poly-AU copolymers of various

composition has enabled the elucidation of the nature of defective sites in the double helix. Adenine (A) is complementary to uracil (U) and pairs with it; uracil does not pair with uracil. The interaction between poly-U and poly-AU may lead to the formation of two structures shown in Fig. 7.9. Either at those sites where U is opposite U

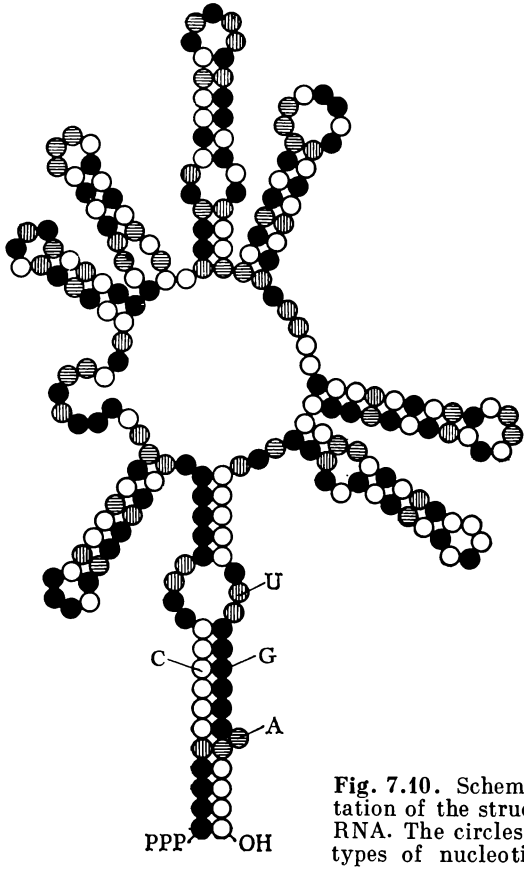


Fig. 7.10. Schematic representation of the structure of phage RNA. The circles show the four types of nucleotides.

here are no hydrogen bonds but the helix retains its helical structure (Fig. 7.9*a*), or the unpaired nucleotides are pushed out of the helix and form *loops* (Fig. 7.9*b*). These two structures differ in the ratio of poly-U to poly-AU with the same degree of polymerization in a system exhibiting maximum hypochromism, i.e., in a double helix. For the structure *a* this ratio is 10 : 10, and for the structure

b it is 6 : 10. Experiment provides an unambiguous evidence in favour of the second possibility—loops are formed.

All types of RNA are characterized by the presence of both extended loops and double-stranded portions, which form "hairpins". Figure 7.10 shows the structure of one RNA. Pairing of bases occurs only in double-stranded portions. The denaturation stability of RNA also increases with increasing content of G-C pairs.

Relatively small molecules of tRNA have been studied especially thoroughly. They also contain double-stranded portions. We shall describe them in Sec. 8.4.

The double helix of DNA is closed into a ring in the cells of bacteria (e.g., in *Escherichia coli*), in particles of bacteriophages, etc. DNA, either circular or unclosed, may be treated on the basis of the rigid-rod model, which is characterized by two energy parameters—axial flexural rigidity and torsional rigidity (see page 83). The persistent length of DNA is close to 60 nm—the double helix has a high flexural rigidity. The circular closed DNA consists of two entangled loops; the order of this entanglement called the linking number Lk is an important characteristic and a topological invariant. Moreover, it is characterized by the number of rotations Mw executed by one of the strands of the double helix relative to the helix axis during motion along the axis. If the axis of the double helix is in the plane, then $Lk = Mw$. In the general case, the nonzero difference of these quantities expresses the degree of twist of the axis of the double helix, $Q = Lk - Mw$. The torsional rigidity of the double helix is determined from the mean-square fluctuation Q, i.e., from $\langle Q^2 \rangle = \langle (\Delta Lk)^2 \rangle - \langle (\Delta Mw)^2 \rangle$. Knowing the value of torsional rigidity, one can obtain an estimate of the free energy required for the alteration of the angle of helical rotation of DNA by a specified amount. For example, in order to increase this angle by 2.6°, which corresponds to the transition of DNA from the *B*- to the *C*-form, only 0.12 RT must be expended. The torsional rigidity of DNA is not high. The theoretical and experimental investigation of the superhelical closed DNA has attracted great attention.

The double-stranded DNA has a specific tertiary structure in solution in the presence of polyethylene glycol. DNA assumes a *compact form of a torus*. Under these conditions CD undergoes a change. This is evidently accompanied by the aggregation of DNA molecules. The diameters of the toruses vary from 80 to 140 nm. The formation of a torus is a reversible conformational transition.

7.3. Intramolecular Interactions in the Double Helix

The structure of the double helix of DNA and double-helical portions of RNA is determined by weak interactions—hydrogen bonds and electrostatic and dispersion forces.

Free nitrogenous bases form hydrogen-bonded complexes in the solid state. The structure of such complexes has been established in a number of cases. There have been found structures that are different from the Watson-Crick model. Figure 7.11 shows the structure of the pair 9-methyladenine-1-methylthymine (MA-MT). The nitrogen atoms N_1 in T and N_9 in A are blocked by methyl groups to avoid the formation of extra hydrogen bonds. We see that the N_1 atom in MT forms a hydrogen bond with the imidazole nitrogen of MA. This structure differs from the Watson-Crick structure. The possibilities of formation of hydrogen bonds between the various atoms of the nitrogenous bases of DNA and also the tautomerism of the nitrogenous bases are essential for *mutagenesis*. A strong mutagen, bromouracil (methylated), has been examined in pairs with ethyladenine and methyladenine. It is interesting that in these two closely related cases there are obtained different structures. In the first pair, hydrogen bonds are formed in O_2—N_6 and N_3—N_7, and in the second, in O_4—N_6 and N_3—N_7.

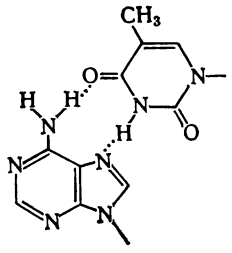


Fig. 7.11. The structure of a 9-methyladenine-1-methylthymine complex.

It is remarkable that in solutions the association of bases occurs predominantly in accordance with the Watson-Crick model. Hydrogen bonds appear in the A-U, A-T, and G-C pairs, but not in the G-U, A-G, and A-C pairs. Accordingly, crystals with noncomplementary pairs cannot be produced from solution.

The melting point of the double helix, i.e., the temperature of the helix-coil transition, depends linearly on the content of G-C pairs. From this it does not, however, follow that the energetics of the double helix is entirely determined by hydrogen bonds. Upon melting the hydrogen bonds between the bases are replaced by bonds with water molecules. Therefore, the difference of hydrogen-bond energies between the helix and the coil is small. There must be other factors that stabilize the double helix.

Aside from the "horizontal" interactions of nitrogenous bases in their planes in the double helix, there also exist *stacking* interactions in the double helix between the bases that belong to neighbouring pairs. These are interactions between parallel and not between coplanar bases directed at right angles to the bases along the axis of the double helix.

Nitrogenous bases are π-electronic systems. Strong dispersion interactions occur between the planar π-electron rings arranged paral-

lel to each other (Fig. 7.12). Therefore, the flat molecules of dyes, whose colour is determined by the conjugation of π-electron bonds, prove to be capable of forming polymers in solution. These polymers, in which the monomers are joined by dispersion forces, have specific optical properties, since they experience an exciton transfer of excitation energy. In the double helices of polynucleotide the parallel packing of bases is manifested, in particular, in the hypochromic effect (see Sec. 5.4).

The first theoretical calculations of the interactions between the nitrogenous bases in DNA demonstrated that the van der Waals

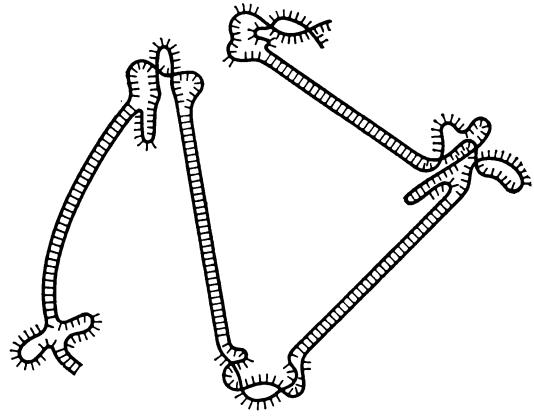


Fig. 7.12. The structure of a polynucleotide in the region of the helix-coil transition.

(i.e., dipole-dipole, induction, and dispersion) interactions in the G-C pairs are considerably stronger than those in the A-T pair. Later, the electrostatic interactions were treated in a more rigorous way, and also the repulsive forces were taken into account with the aid of atom-atom potentials. A method was devised for calculation of the energies of horizontal and stacking interactions which are computed as the sums of atomic interactions. What is determined here is the sum of the energies of electrostatic and polarizing interactions and the energy of repulsion. The charges on the atoms and bonds are found by means of approximate methods of quantum chemistry. The efficiency of the method of calculation is tested by calculations of the energies of a number of molecular crystals and comparison of these with the experimental values of the heats of sublimation. The values found for naphthalene and anthracene are, respectively, 75.6 kJ/mole (the experimental value is 72.4) and 104.2 (experiment, 98.8, 100.0, and 104.2). These data point to the order of magnitude of stacking. The greatest energy corresponds to two

neighbouring antiparallel C-G and G-C pairs. We shall not cite numerical data, since they are as yet not unambiguous and depend substantially on the method of calculation. Presumably, the energy of stacking interactions varies for different pairs between 40 and 60 kJ/mole, which appreciably exceeds the hydrogen-bond energy.

7.4. Thermodynamics of Melting of the Double Helix

The denaturation of nucleic acids boils down to the destruction of the double helix (DNA) or helical double-stranded portions (RNA). The heating of a solution of native DNA causes the separation of the double helix into two chains, which form random coils. In this process, the viscosity and optical activity are diminished and the hypochromism disappears, i.e., the absorption intensity in the 260 nm region increases. The separation into two chains is directly proved by the centrifugation of DNA containing ^{15}N in the CsCl density gradient (cf. page 88). The cells of *E. coli* grown in a culture medium on ^{15}N were transferred into a medium with ordinary ^{14}N. Upon cell division there were formed replicated double helices, in which one chain contained ^{15}N and the other ^{14}N. Before denaturation there was observed one density peak (1.717 g/cm^3) corresponding to ^{15}N-^{14}N double helices. After denaturation two peaks appear—1.740 and 1.724 g/cm^3 corresponding to single-stranded coils with ^{15}N and ^{14}N, respectively. The density increases, since the coils are more compact than the helix. The molecular weight of DNA diminishes twice upon denaturation. The formation of coils is observed in the electron microscope.

The simplest model of the double helix is formed by two complementary chains of homopolynucleotides, say, poly(A) and poly(U). This double helix melts at 65°C in a 0.15*M* solution of NaCl at pH 7.0. The intensity of the 260-nm absorption band increases on melting by 34 percent and the specific rotation φ_{sp} diminishes by 275°.

The linear dependence of the melting temperature of DNA on the content of G-C pairs gives on extrapolation the limiting maximum values of T_m = 69°C for poly (A-T) and 110°C for poly (G-C), which agree well with experimental values for the corresponding synthetic polynucleotides (65 and 104°C). The melting point of DNA increases with increasing ionic strength of the solution approximately in proportion to the logarithm of the cation concentration. This increase is understandable—the higher the cation concentration, the more compensated are the negative charges of the phosphate groups and, hence, the weaker is the repulsion between the complementary chains.

The melting of DNA has been studied by Doty as back as 1957. This phenomenon is specific; DNA is a unique example of a one-dimensional ordered structure with an aperiodic distribution of units.

Before turning directly to DNA it will be useful to consider a simpler double helix built up of two complementary homopolynucleotide chains. In the region of the helix-coil transition such a molecule consists of alternating helical and disordered portions—loops (Fig. 7.12). If we denote the number of disrupted pairs by N_1 and the number of bound pairs by N_2 and the number of double-stranded portions is assumed to be equal to the number of loops n, then the free energy of the system will be written in the form

$$G(N_1, N_2, n) = N_1G_1 + N_2G_2 + nG_s - TS_0 \qquad (7.1)$$

where G_1 and G_2 are the free energies of the completely separated and fully double-stranded molecules referred to one base pair; G_s is the free energy required for the formation of a loop, i.e., the molten region between two helical regions. Finally, S_0 is the entropy of mixing of the helical and nonhelical portions, i.e. (per mole):

$$S_0 = R \ln \frac{N_1!}{n!\,(N_1-n)!} \frac{N_2!}{n!\,(N_2-n)!} \qquad (7.2)$$

The minimum of G at a given temperature corresponds to the condition

$$\left(\frac{N_1}{n} - 1\right)\left(\frac{N_2}{n} - 1\right) = \frac{1}{\sigma} \qquad (7.3)$$

where $\sigma = \exp(-G_s/RT)$ is the cooperativity factor (see page 110). The lower the value of σ, the higher is the cooperativity. The equilibrium values of N_1, N_2, and n are found by differentiation of G with respect to N_1, the value of N_2 being equal to the difference $N - N_1$, where N is a constant. Equating $\partial G/\partial N_1$ to zero, we obtain:

$$\frac{1-n/N_2}{1-n/N_1} = s = \exp\frac{G_1-G_2}{RT} \qquad (7.4)$$

Thus, a homopolymer in the transient region is a sequence of helical and nonhelical portions, whose dimensions are determined by σ. This is associated with the unidimensionality of the system, which cannot be separated into phases according to the *Landau-Lifshitz theorem* (1930).

The melting curve, i.e., the dependence of the fraction of disordered pairs $N_1/N = 1 - \theta$ (where θ is the degree of helicity) on temperature T, is the steeper, the smaller is the value of σ. At $\sigma = 1$ no cooperativity is present and at $\sigma = 0$ the cooperativity is complete. In the midpoint of the melting range the derivative $|d\theta/dT|$ is maximal and the range is determined by the following condition:

$$\Delta T = \frac{1}{|d\theta/dT|_{\max}} \qquad (7.5)$$

The course of the curve is shown in Fig. 7.13. A calculation based on the Ising model (see pages 79, 111) yields:

$$\Delta T = 4\sigma^{1/2} \frac{RT_m^2}{\Delta H} \tag{7.6}$$

where ΔH is the enthalpy change between the helical and the nonhelical molecule per base pair. From the experimental values of ΔT for synthetic homopolynucleotides there is obtained $\sigma \sim 10^{-4}$ to 10^{-5}, i.e., $G_s \approx 30$ kJ/mole. The extent of cooperativity is very high.

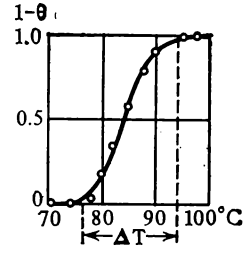


Fig. 7.13. The helix-coil transition in DNA.

The detailed studies of heteropolymers carried out by Lazurkin, Frank-Kamenetsky, and their colleagues have made it possible to build a theory applicable to DNA and its complexes with small molecules. The treatment of heterogeneity requires that two factors be taken into account—the difference in stability between the A-T and G-C pairs and an extra strengthening or weakening of the double helix by ligands. The influence of these factors is different, since the ligands added can be redistributed along the chain during melting and the primary structure is not altered.

Assume that m_1 molecules of the ligand are linked together by nonhelical portions of the polymer and m_2 molecules by the helical portions. In place of Eq. (7.1) we have

$$G(N_1, N_2, n, m_1, m_2) = N_1G_1 + N_2G_2 + nG_s + \\ + m_1g_1 + m_2g_2 - TS_0 - RT \ln \frac{N_1!}{m_1!\,(N_1-m_1)!} \frac{N_2!}{m_2!\,(N_2-m_2)!} \tag{7.7}$$

where g_1 and g_2 are the free energies of the ligands in the nonhelical and helical portions, respectively. The last term incorporates the entropy of mixing of the polymer with the ligands. The minimum condition $\partial G/\partial n = 0$ coincides with Eq. (7.3), i.e., the average length of the helical sections at a given value of θ does not depend on the presence of ligands. On the other hand, from the condition $\partial G/\partial N_1 = 0$ we have instead of Eq. (7.4):

$$\frac{1-n/N_2}{1-n/N_1} = s\frac{1-c_1}{1-c_2} = s^* \tag{7.8}$$

where $c_1 = m_1/N_1$ and $c_2 = m_2/N_2$. These ligand concentrations are expressed in terms of the ligand concentration in solution, c_0,

and the binding constants K_1 and K_2 in the following manner:

$$c_1 = \frac{c_0}{1/K_1 + c_0} \quad \text{and} \quad c_2 = \frac{c_0}{1/K_2 + c_0} \tag{7.9}$$

The melting curve for a homopolymer without ligands is described by the following function:

$$\theta = f(s) \tag{7.10}$$

with θ being equal to 0.5 at $s_m = 1$, i.e., at $G_1 = G_2$. In the presence of ligands

$$\theta = f(s^*) \tag{7.11}$$

and $\theta = 0.5$ at $s^* = 1$, i.e., at (see Eq. 7.8)

$$s_m = \frac{1-c_2}{1-c_1} \tag{7.12}$$

The shift of the melting temperature as compared with the temperature shift for a pure polymer T_0 is given by

$$\delta\left(\frac{1}{T_m}\right) = \frac{1}{T_0} - \frac{1}{T_m} = \frac{R}{\Delta H} \ln \frac{1-c_1}{1-c_2} \tag{7.13}$$

The change of the melting range is

$$\delta\left(\frac{\Delta T}{T_m^2}\right) = \frac{\Delta T}{T_m^2} - \frac{\Delta_0 T}{T_0^2} = -\frac{R}{\Delta H}\left[\frac{\partial}{\partial\theta} \ln \frac{1-c_1}{1-c_2}\right]_{\theta=0.5} \tag{7.14}$$

The total ligand concentration in solution and on the polynucleotide is equal to

$$c = c_0 + (1-\theta)\frac{c_1 p}{2} + \theta \frac{c_2 p}{2} \tag{7.15}$$

where p is the concentration of the phosphate groups of the polynucleotide that bind the ligands. If $c \gg p$, then

$$\delta\left(\frac{1}{T_m}\right) = \frac{R}{\Delta H} \ln\left(\frac{1+K_2 c}{1+K_1 c}\right) \tag{7.16}$$

and

$$\delta\left(\frac{\Delta T}{T_m^2}\right) = 0 \tag{7.17}$$

Conversely, on strong binding of ligands by the polymer over the entire transition region, i.e., at $K_1 p \gg 1$ or $K_2 p \gg 1$ and $c \approx p$:

$$\delta T_m = T_m - T_0 = 2\left(\frac{q-1}{q+1}\right)\frac{RT_0^2}{\Delta H}\frac{2c}{p} \tag{7.18}$$

$$\delta\Delta T = \Delta T - \Delta_0 T = 4\left(\frac{q-1}{q+1}\right)^2 \frac{RT_0^2}{\Delta H}\frac{2c}{p} \tag{7.19}$$

where $q = K_2/K_1$.

Thus, ligands act as "clips" that stabilize the double helix.

Experiment confirms the above relations. Equation (7.16) describes, for instance, the effect of pH on melting curves. It has been used to find the values of $\Delta H = 40$-45 kJ/mole at high ionic strengths and neutral pH. The ligands studied were acridine dyes, the ions of alkali metals, Ag^+, Cu^+, etc. The ions Ag^+ and Cu^+ are predominantly bound by G-C pairs in the DNA double helix. Investigations of the binding of metal ions gave the values of ΔH equal to $34 \pm \pm 4$ kJ/mole at $0.01M$ NaCl and to 46 ± 8 kJ/mole at $0.1M$ NaCl. The above determinations of ΔH are in agreement with the results of direct calorimetric measurements.

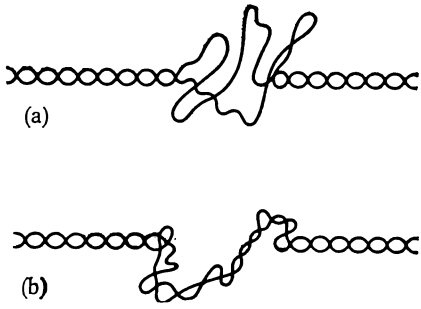


Fig. 7.14. The melting of the double helix:
(*a*) with formation of loops; (*b*) without formation of loops.

We shall now turn to the melting of a helical double-stranded heteropolymer. Such melting can occur with or without formation of large loops (Fig. 7.14). A theory has been worked out for both cases. With a disordered distribution of A-T and G-C pairs there is obtained a linear dependence of T_m on the content of G-C pairs (cf. page 250).

A special feature of a heteropolymer is the relatively small number of microstates corresponding to a particular energy. It is unlikely that two different distributions of coils and helices along the chain with the same values of N_1, N_2, and n will have the same energy, since they will undoubtedly contain different numbers of G-C pairs in molten portions. The entropy of mixing cannot therefore have a substantial effect on the melting of heteropolymers. At the same time, there appears a new energy factor, which is determined by the fact that when the average length of molten portions decreases, the content of stable G-C pairs in them must diminish. The competition between this factor and the factor due to the unfavourableness of the "joints" between the helical and nonhelical portions must lead to the alternation of helical and nonhelical regions of definite length at a given value of θ.

An approximate solution of the problem of the melting of a heteropolymer with a disordered sequence of base pairs has been arrived at by Lazurkin and Frank-Kamenetsky. Let us split the molecule into identical sections, each containing λ pairs. If λ is sufficiently large, then the distribution of the concentration of G-C pairs over

these sections will be a Gaussian distribution, i.e.,

$$P_\lambda(x) = \frac{1}{\sqrt{2\pi}\, b_\lambda} \exp\left[-\frac{(x-x_0)^2}{2b_\lambda^2}\right] \tag{7.20}$$

where $b_\lambda^2 = x_0(1-x_0)/\lambda$; x is the content of G-C pairs; x_0 is the value of x corresponding to the maximum of the distribution.

The total number of G-C pairs in molten portions is a minimum if all portions with a content of such pairs less than a certain limiting value x_λ, are molten and all sections with $x > x_\lambda$ are helical. The value of x_λ is determined by the condition

$$\int_0^{x_\lambda} P_\lambda(x)\, dx = \frac{N_1}{N} \tag{7.21}$$

The mean concentration of G-C in molten sections is equal to

$$\bar{x}_1 = \int_0^{x_\lambda} xP_\lambda(x)\, dx \left(\int_0^{x_\lambda} P_\lambda(x)\, dx\right)^{-1} = \frac{N}{N_1}\int_0^{x_\lambda} xP_\lambda(x)\, dx \tag{7.22}$$

Since the composition of each section is independent of the composition of the other sections, the probability of a molten region being composed of r successive sections is given by

$$w_r = (1-N_1/N)(N_1/N)^{r-1} \tag{7.23}$$

and the average number of sections in the molten region is

$$\bar{r} = (1-N_1/N)^{-1} \tag{7.24}$$

The average number of base pairs in the molten region is equal to

$$\bar{m}_1 = \lambda\bar{r} = \lambda(1-N_1/N)^{-1} \tag{7.25}$$

and the number of molten regions in the entire molecule is given by

$$n = N_1/\lambda\bar{r} = N_1(1-N_1/N)/\lambda \tag{7.26}$$

The free energy of the polymer is

$$G(N_1, N_2, n) = N_1[\bar{x}_1 G_1^{GC} + (1-\bar{x}_1)G_1^{AT}] + N_2[\bar{x}_2 G_2^{GC} + (1-\bar{x}_2)G_2^{AT}] + nG_s \tag{7.27}$$

where G_s is the entropy contribution determined by the mixing of regions. Introducing the conditions $N_1 + N_2 = N$ and $\bar{x}_1 N_1 + \bar{x}_2 N_2 = x_0 N$, expressing n in terms of λ according to Eq. (7.26), using the notations $\Delta G_{GC} = G_1^{GC} - G_2^{GC}$ and $\Delta G_{AT} = G_1^{AT} - G_2^{AT}$

and discarding the constant terms, we obtain

$$G(N_1, \lambda) = N_1 \bar{x}_1 \Delta G_{GC} + N_1 (1 - \bar{x}_1) \Delta G_{AT} + N_1 (1 - N_1/N) G_s/\lambda \tag{7.28}$$

Here $\bar{x}_1$ is defined by formula (7.22) with x_λ from Eq. (7.21). The equilibrium values of N_1 and λ are found from the minimum condition for the expression $G(N_1, \lambda)$ with respect to both variables. In order to find the average length of the helical region and the melting range of the heteropolymer ΔT, it is necessary to consider expressions (7.21), (7.22), and (7.28) near $N_1 = 1/2\ N$, i.e., to carry out an expansion into a Taylor series in $z = N_1/N - 1/2$. Using Eq. (7.20) and again dropping the constant terms, we arrive at the following expression:

$$G(N_1, \lambda) = N \left\{ z\Delta G - \frac{1}{\sqrt{2\pi}} \left[\frac{x_0 (1-x_0)}{\lambda} \right]^{1/2} (\Delta G_{GC} - \Delta G_{AT}) \times \right.$$
$$\left. \times (1 - \pi z^2) + \frac{G_s}{4\lambda} (1 - 4z^2) \right\} \tag{7.29}$$

where

$$\Delta G = x_0 \Delta G_{GC} + (1 - x_0)\ \Delta G_{AT} \tag{7.30}$$

From the condition $\partial G/\partial \lambda = 0$ we obtain at $z = 0$

$$\lambda_0 = \frac{\pi}{2} \left(\frac{G_s}{\Delta G_{GC} - \Delta G_{AT}} \right)^2 \frac{1}{x_0 (1-x_0)} \tag{7.31}$$

where we have

$$\Delta G_{GC} - \Delta G_{AT} = (T_{GC} - T_{AT})\ \Delta H_{AT}/T_{AT} \tag{7.32}$$

where T_{GC} and T_{AT} are the melting points of the corresponding polymers. Using Eq. (7.25), we find that

$$\bar{m}_1 = \pi \left[\frac{T_{AT}}{\Delta H_{AT}} \frac{G_s}{T_{GC} - T_{AT}} \right]^2 \frac{1}{x_0 (1-x_0)} \tag{7.33}$$

Equating to zero the derivative of the free energy with respect to z at constant λ and differentiating the resulting expression with respect to T, we find

$$\Delta T = 2 \frac{\pi - 2}{\pi} \frac{\Delta H_{AT}}{T_{AT}} \frac{(T_{GC} - T_{AT})^2}{G_s} x_0 (1 - x_0) \tag{7.34}$$

A rigorous solution of the problem differs from the above solution only in numerical factors. The formation of loops on melting is taken into account by an additional contribution to G corresponding to the entropy of the loops.

Numerical calculations on electronic computers with the use of the fraction of G-C pairs x_0, the melting points of homopolymers, T_{AT} and T_{GC}, and also ΔH_{AT} and σ as independent parameters have made it possible to determine the melting curves, the dependence of

the average length of the helical section on the degree of helicity θ, and the dependence of the transition range on x_0 at different values of the parameters indicated above. Figures 7.15 through 7.17 show the curves plotted at $T_{AT} = 340$ K, $T_{GC} = 380$ K, $\Delta H_{AT} = 29$ kJ/mole, $\sigma = 5 \times 10^{-5}$, $N = 2 \times 10^4$ to 10^6. A linear plot of T_m vs. x_0 is obtained directly (in K):

$$T_m = T_{AT} + (T_{GC} - T_{AT})\, x_0 = 340 + 40x_0 \qquad (7.35)$$

Taking account of the formation of loops diminishes somewhat the value of ΔT. The shape of the theoretical melting curve is in accord

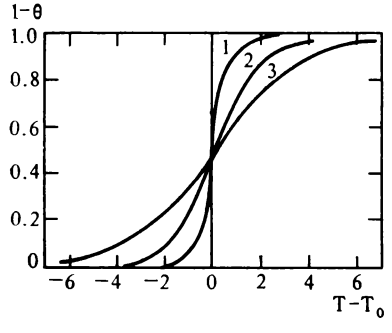


Fig. 7.15. The temperature dependence of 1—θ for polyheteronucleotide: *1*—$x_0 = 0$, $T_m = 340$ K, $N = 2 \times 10^4$; *2*—$x_0 = 0.1$, $T_m = 344$ K, $N = 4 \times 10^5$; *3* — $x_0 = 0.5$, $T_m = 360$ K, $N = 6 \times 10^5$.

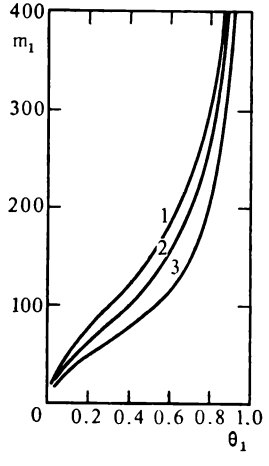


Fig. 7.16. The dependence of the average number of base pairs in molten regions on $\theta_1 = 1 - \theta$. The curves *1*, *2*, and *3* correspond to the same parameters as in Fig. 7.15.

with experiment, practically irrespective of the model. Conversely, the theoretical graph for ΔT depends substantially on the model adopted and the values of the parameters chosen. All these parameters, except σ, are found from independent experiments. An excellent agreement with experiment has been obtained for a number of DNAs at $\sigma = 5 \times 10^{-5}$. Thus, experiment allows one to find σ. The cooperativity of the melting process is very high.

Determinations of the optical activity (and, hence, of the degree of helicity θ) and dimensions of the DNA molecule in solution by measuring intrinsic viscosity during melting have made it possible to find the average lengths of helical portions. In accordance with

theory, in the region of $\theta = 0.8$-0.9 the lengths of helical portions vary from 1000 to 2500 pairs, whereas in the transition region, i.e., at $\theta = 0.5$ we have $m = 400$-500 pairs.

The melting point decreases and ΔT increases with decreasing chain length.

The thermodynamic characteristics of melting of DNA have been studied by the methods of microcalorimetry (Privalov). The heat of melting and the melting point strongly depend on pH. As the pH value increases from 7.0 to 9.7 the melting point T_m falls from 84.8 to 66.3°C, ΔH from 40.4 to 29.9 kJ/mole, and ΔS from 113 to 88 J/(mole·K). As the pH decreases from 5.4 to 3.2 the melting point drops from 84 to 55°C, ΔH from 39.5 to 10.4 kJ/mole, and ΔS from 107 to 52 J/(mole·K). The value of T_m depends on ionic strength.

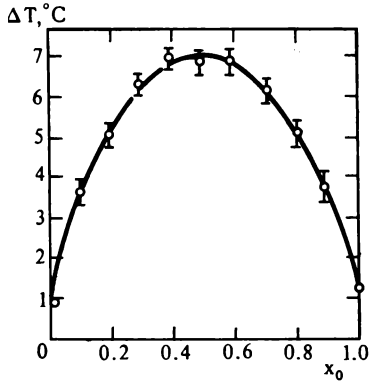


Fig. 7.17. The melting temperature range versus the concentration of guanine cytosine (G-C) pairs.

The main results of theoretical calculations for real DNA molecules with a random nucleotide sequence under physiological conditions are summarized by Lazurkin as follows.

1. The width of the melting range ΔT is 2.5-3°C.

2. The dependence of the average length of a helical portion on the helicity θ is depicted by a sigmoidal curve.

3. Circular closed DNAs that are present in phages and bacteria must melt at a higher temperature than do the same DNAs in the linear form because of the topological impossibility of free unwinding with the formation of loops.

Experiment confirms these predictions for a number of bacteriophages. At the same time, the appreciably greater values of ΔT for the DNA of bacteria (about 5°C) and higher organisms (about 10°C) led to the conclusion that these DNAs have a block structure. The entire DNA in genome may be split into a number of sequences which will differ in the average content of G-C pairs and each of which will have a quasi-random base sequence. The molecular mass of blocks is equal to $(5\text{-}15) \times 10^9$.

The increased precision of optical measurements allowed the detection of the fine structure of melting curves. It was first observed for the phage T7 DNA. To reveal the fine structure it is convenient to depict melting curves in the form of the temperature dependence of

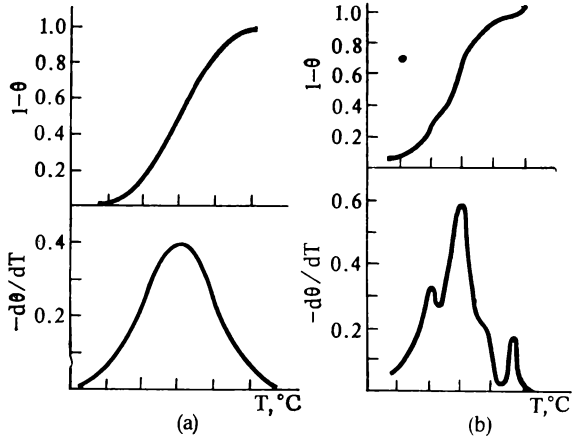


Fig. 7.18. Ordinary and differential melting curves for DNA: (*a*) smooth curves; (*b*) curves with a fine structure.

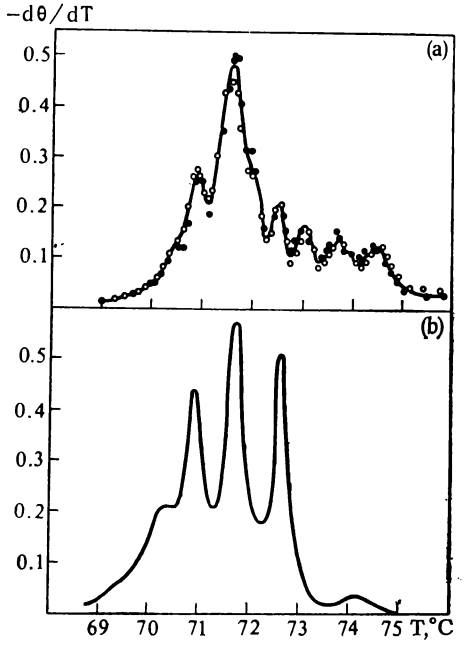


Fig. 7.19. The differential melting curves for the replicative double-stranded form of DNA from phage φX174.

the derivative of the degree of melting $(1 - \theta)$ with respect to temperature. This is shown in Fig. 7.18. Use is made of the methods of graphical or computer differentiation of melting curves or automatic differentiation of melting curves during measurement. The latter is accomplished, in particular, by measuring the difference in light absorption in two cells, whose temperature is maintained identical but a small difference of T_m is created between the solutions because of the difference in salt concentration. The fine structure of the differential melting curve (DMC) is observable for not too long double helices of DNA; it disappears if the number of pairs exceeds 10^5.

Theory shows that to each DMC peak there corresponds the melting of a relatively long portion, whose content of G-C pairs is such that its T_m coincides with the temperature of the experiment. Theory explains the details of the DMC. Figure 7.19*a* shows the experimental DMC of the phage ϕX174 and Fig. 7.19*b* gives the theoretical DMC calculated on the basis of the known sequence of 5375 nucleotides in this DNA.

7.5. Kinetics of Unwinding of the Double Helix

We have not yet examined the problems of the rates of helix-coil transitions. These problems are not trivial.

The kinetics of the unwinding of the DNA double helix was first studied by Kuhn (1957). Unwinding occurs after the interchain bonds are disrupted. If we assume that it takes place as a result of rotational Brownian motion, then the time τ required for this process to take place will be much longer than that observed. For instance, for DNA with a molecular weight of 3×10^6 the unwinding of 450 turns of the helix, which is necessary for complete separation, will require 150 days. Meanwhile, the time τ for DNA with a molecular weight of the order of 10^8 is about 1 min. Kuhn examined the separation of chains that occurs when rotational thermal motion is combined with translational motion and obtained the time τ of the order of 1 min for DNA with a molecular weight of 3×10^6, which is also too long.

The unwinding time is considerably shortened if we take into account the rotational moment that arises as a result of an increase in entropy for separate chains (Longuet-Higgins and Zimm, 1960). This moment is equal to

$$P = \frac{T\Delta S}{\varphi} = \frac{57.3}{36} T\Delta S \tag{7.36}$$

where $\varphi = 36/57.3$ is the angle expressed in radians by which the chain end must be rotated in order to release one base pair. The rate of unwinding, i.e., the relative angular velocity of the two halves of the helix, is $\omega = P/\beta$, where β is the coefficient of friction. For helices with a molecular weight of the order of 10^6 the time τ turns out to be of the order of seconds. This estimate does not take

into account the irreversibility of unwinding and entanglement of the released chains.

Without dwelling on other models, we turn to the theory proposed by Fong (1964). The entanglement of free chains must not occur if the interchain bonds are broken at the midpoint of the double helix. The two halves of the helix fluctuate independently, half the time spent on unwinding and the other half on rewinding. However, at the initial stage of strand separation, because of the tight packing of the helix, the second winding cannot occur and the kinetic energy of entangled chains is converted to the potential energy which contributes to unwinding. If many turns have occurred, the chains may be twisted back again. Thus, the two halves of the helix are randomly twisted and untwisted with the helix radius remaining unchanged. But twisting cannot continue for an indefinitely long time period, and therefore with such random motion the helix is predominantly unwound.

The number of rotations $N(t)$ on unwinding is less than or equal to the initial number of turns of the helix N_0. We must find τ from the condition

$$\int_0^{\tau} \omega(t)\, dt = -2\pi N_0 \tag{7.37}$$

where $\omega(t)$ is the relative angular velocity of the two halves of the helix. If $N(t) = N_0$, then the number of unwinding turns is equal to the number of rewinding turns, $\omega(t_-) = -\omega(t_+)$. We can evaluate the upper and lower limits of τ. The average value of $\omega(t)$ over a certain time interval $\bar{\omega}(t)$ is a monotonously decreasing function of t, since more than the doubled time period is required for the unwinding of the double helix with the doubled number of turns. The lower limit of τ is obtained upon replacement of $\bar{\omega}(t)$ with the initial maximum value of $\bar{\omega}(0)$. Let the initial relative angular velocity be ω_0. If the two halves of the helix unwind, then $\omega(0) = 2\omega_0$. Since at time $t = 0$ the DNA helix is unwinding during half the time and winding during the other half, then $\bar{\omega}(0) = 1/2\omega(0) = \omega_0$. This value is found from the condition

$$1/2 I\omega_0^2 = 1/2\ kT \tag{7.38}$$

where I is the moment of inertia of one half of the helix relative to its axis. At $N_0 = 2 \times 10^4$ (the molecular weight 1.2×10^8, the number of base pairs 2×10^5) and $T = 20°C$, the value of ω_0 is equal to 3×10^8 rad/s. The lower limit of τ is given by

$$\tau_{\text{lower}} = 2\pi N_0/\bar{\omega}(0) = 4.2 \times 10^{-4}\ \text{s}$$

The upper limit τ_{upper} corresponds to the random winding and unwinding without the latter predominating. We assume that the

DNA double helix is in water at 20°C. The mean rotational moment P of one half of the molecule having a molecular weight of 1.2×10^8 is 1.34×10^{-22} esu and the momentum of the water molecule is $p = 1.9 \times 10^{-18}$ esu. Multiplying p by the helix radius, we find the mean rotational moment that arises as a result of the collision of the molecules of H_2O and DNA, $p' = 1.7 \times 10^{-25}$ esu. Thus, $P/p' = q = 7.8 \times 10^2$. To reduce P to zero, a large number of collisions (equal to q^2) is required. The calculation gives 2×10^{15} collisions per second and the time required for q^2 collisions, i.e., for the moment to be reduced from P to zero, is 3×10^{-10} s. The average time $\bar{t}$ between two successive zeros of the fluctuating moment is twice as long due to the symmetry of fluctuations in time. During the time $\bar{t}$ the angular velocity does not change sign and the time $\bar{t}$ and the number of rotations n during this time may be considered to be a step in Brownian wandering. We have $n = \omega_0 \bar{t} = 2.9 \times 10^{-2}$ turns and the time τ_{upper} for 10^4 turns is equal to $(10^4/n)^2\, \bar{t} = 73$ s. At $N \gg 1$ the time τ is close to the upper limit. This value is in agreement with experiment. From the calculation it follows that τ_{upper} is proportional to M^3 and τ_{lower} is proportional to $M^{3/2}$ (M is the molecular weight).

Experiment shows that τ is proportional to M^2 for a number of phage DNAs. The quadratic dependence follows directly from the assumption that τ must be proportional to the number of turns, i.e., M, and the viscosity of the medium, i.e., again the molecular weight. Massey and Zimm (1969) studied the denaturation of DNA by means of relaxation methods and established the dependence of τ on a number of factors—the viscosity and ionic strength of the medium, the concentration of DNA, its molecular weight, the stage of helix-coil transition. We have

$$\tau = b\,(1 + c\,[\eta])\,\eta_0 M\psi \qquad (7.39)$$

where $0 < \psi < 1$ is the factor which characterizes the transition step; c is the concentration of DNA; $[\eta]$ is its intrinsic viscosity; η_0 is the viscosity of the solvent; b is a constant which depends on the ionic strength. At small c the value of τ is proportional to M. The increase of τ from the start of the transition to the end shows that the resistance of the medium increases as the molecule is unwound. As a result of the formation of loops, the effective radius of the double helix increases (the Fong theory does not take it into account).

In actual fact, the kinetics of the melting of DNA corresponds to an entire spectrum of times τ. The true theory of this phenomenon has not yet been worked out: serious difficulties are encountered in the investigation of the relation between the attainment of internal equilibrium and unwinding. The disordered sections of the helix may undergo "fusion" again.

Varshavsky and Evdokimov have studied the unwinding of DNA

by the thermal impact method. A solution of DNA was heated from 5 to 20°C for 0.5 s. The kinetic curves indicate the presence of at least two steps of the structural transition. In the first, rapid step almost all the hypochromic effect disappears. Presumably, at this step there are formed fixed loops, which are fully unwound at the second stage. The rate of unwinding obeys the Arrhenius equation (page 188), and the energy of activation strongly depends on pH, attaining a maximum value at intermediate pH values.

An experimental investigation of the kinetics of unwinding enables one to determine defects in the secondary structure of DNA. For this purpose, Lazurkin and his coworkers have worked out the *kinetic formaldehyde method* (KFM).

If the solution contains a substance which reacts with partially despiralized nucleotides and which thereby hinders the formation of Watson-Crick pairs by these nucleotides, the reaction will then proceed until complete despiralization is attained. If the time of binding of such a reagent exceeds the unwinding time, then this reaction is exactly the one that limits the rate of unwinding. The rate constant of the reaction of native DNA with the reagent is $k = k'w$, where w is the probability of separation for any pair of nucleotides in the molecule and k' is the rate constant of the reaction of the separated nucleotides with the reagent. If some of the bases of native DNA have reacted and formed a locally denatured portion, the probability of the thermal denaturation of the adjacent pairs will considerably increase, $w' \gg w$. The rate constant k' is accordingly much greater than k. The formation of a locally denatured segment requires a free energy G_s (see page 251). Therefore, $k/k' = w/w' \approx \exp(-G_s/RT) = \sigma \sim 10^{-4}$ to 10^{-5}. Since $k \ll k'$, the unwinding process, which at first proceeds slowly, is gradually accelerated and the curve showing the dependence of $1 - \theta$ on time is a sigmoidal one.

If the double helix has imperfect regions—originally despiralized sections, the initial slope of the curve is increased. The curve is shown in Fig. 7.20. Theory yields the following formula:

$$-\ln \theta = (ky + 2k'c)\, t + kk't^2 \qquad (7.40)$$

where y is the initial dimensions of the unwinding centre expressed in terms of the number of base pairs; $c = \bar{n}^{-1}$ is the concentration of defects; $\bar{n}$ is the average number of base pairs in the helical portion. The dependence of $-\ln \theta/t$ on t is a straight line which makes an intercept on the ordinate:

$$z = ky + 2k'c \qquad (7.41)$$

Formaldehyde is such an unwinding agent. Let us evaluate the right-hand side of formula (7.41). From experiments with mononucleotides it follows that the rate constant k' of the reaction between a base and formaldehyde is of the order of 1 min^{-1}. Hence, $k = wk' \sim$

$\sim \sigma k' \sim 10^{-4}$ to 10^{-5} min^{-1}. The value of y is unknown, but it is reasonable to put it equal to a few units. Thus, $ky \sim 10^{-4}$ min^{-1}. This value lies below the sensitivity limit of kinetic experiments and may be neglected as compared with $2k'c$, since $k' \gg k$ and c is of the order of units. Thus,

$$z = 2k'c \tag{7.42}$$

The constant k' is found from experiments on fragmented DNA from formula (7.42)—in such cases c is unknown. The order of mag-

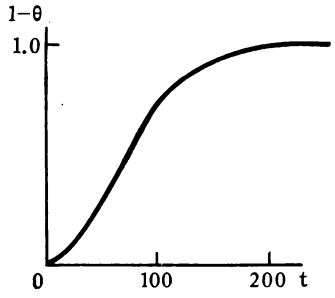


Fig. 7.20. A kinetic curve for DNA with defects (t in relative units).

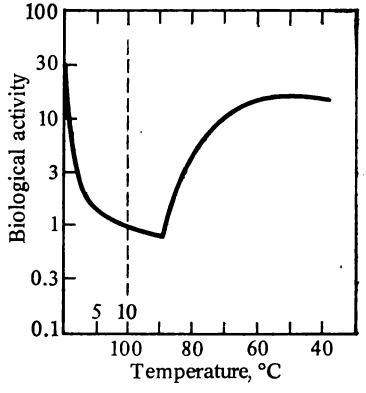


Fig. 7.21. Renaturation of DNA. The numbers 5 and 10 above the abscissa indicate the time during which the samples are kept at 100°C.

nitude of k' is 2.5 min^{-1}. The lowest of z that can be measured is 3×10^{-4} min^{-1}.

In other words, the kinetic formaldehyde method (KFM) allows us to determine the minimum concentration of defects c of the order of one defect per 10^4 nucleotide pairs. The method is very sensitive. It is used with success for studying DNA with breaks caused by ultraviolet radiation or enzymatic hydrolysis, etc. Native DNA is found to be practically free of defects, which disproves the zig-zag model with sharp breaks. KFM is applicable to the study of complexes of DNA with RNA-polymerase and of the fine structure of DNA.

The denatured coil-like single-stranded chains of DNA are capable of renaturation: they can again form a double helix. The *renaturation* of DNA was discovered and studied in 1961 by Marmur and Doty. Slow cooling, i.e., annealing, of denatured DNA results in a partial recovery of the biological activity (Fig. 7.21) and hypochromic

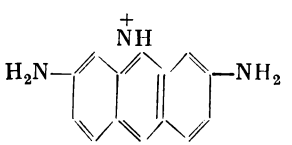
effect. Renaturation also occurs after rapid cooling (quenching) of denatured DNA if the sample is again subjected to heating up to a temperature somewhat lower than T_m. Renaturation is easily followed from a change in density with the aid of sedimentation in the CsCl density gradient (see pages 88 and 250). For instance, the density of native DNA is 1.704 g/cm^3, whereas the denatured and quenched (in this state) native DNA has a density of 1.716 g/cm^3 and the density of renatured DNA is 1.700 g/cm^3. Renaturation is greatly dependent on temperature; the optimum temperature for renaturation is the temperature high enough for unwinding of coils and is at the same time low enough for the double helix to be stabilized. DNA from *Diplococcus pneumoniae* undergoes renaturation at 67°C at a rate 2.5 times as faster as at 80 and 50°C.

Renaturation proceeds readily in the case of phase DNA, with some difficulty for bacterial DNA and is practically infeasible for the DNA of higher organisms. This is understandable—if the DNA samples produced from calf thymus, cells of *E. coli* and phage T2 had all a molecular weight of 10^7, the number of various DNA molecules per cell (or phage particle) would be 10^6, 10^2, and 10, respectively. With the same weight concentration all the three samples would have quite different concentrations of complementary chains, and the probability of an encounter of two such chains formed on denaturation of DNA in the sample produced from the cells of a higher organism would be very low.

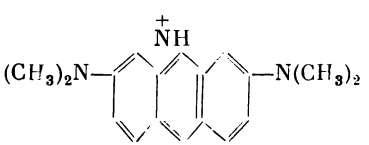
Renaturation can lead to the formation of hybrid duplexes from DNA's of various origin. There have been obtained hybrids of DNA isolated from various strains of *E. coli*, and also interspecies bacterial hybrids. Hybridization of DNA discloses the evolution-genetic relations between bacteria.

7.6. Interaction of the Double Helix with Small Molecules and Ions

The study of the interaction of DNA and RNA with small molecules is important for the elucidation of the structure of nucleic acids and its transformations. Small molecules have in a number of cases a significant effect on the biological function of DNA and RNA. Some of them are mutagens, others inhibit the transcription (see Sec. 8.8). Mutagens include, for example, *acridinic dyes*:



Proflavin



Acridine orange

On the other hand, protein molecules, which are also small as compared with DNA, regulate, while being linked to DNA, the basic molecular-genetic processes.

The theory of adsorption of small molecules on DNA has been developed by Crothers and Gursky and Zasedatelev.

Suppose that a biologically active ligand is bound by a certain number of base pairs of DNA. Let us consider the equilibrium binding of ligands by a polymer, which is a random sequence of A-T and G-C pairs. The ligand may contain functional groups, part of which undergo specific interaction with base pairs. Let α be the ligand length

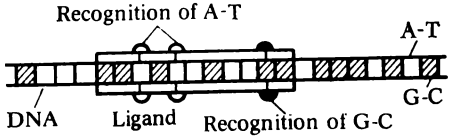


Fig. 7.22. Recognition of the DNA double helix by the ligand molecule.

and q_{AT} and q_{GC} the numbers of active sites per ligand molecule which recognize A-T and G-C pairs. Each site is bound to one pair.

Let us assume that the ligand is adsorbed by the polymer if all the sites of the latter are bound with the corresponding base pairs. The scheme of such binding is shown in Fig. 7.22. The values of α, q_{AT}, and q_{GC} and the binding constant K can be determined from the experimental binding curve by comparing it with the theoretical isotherm calculated with the aid of a computer. A simpler procedure consists of a theoretical calculation of the effective binding constants K_{eff}, the number of binding sites n_{eff}, and the saturation level for binding, r_{max}. The calculated values are compared with experimental values with the aid of a Scatchard graph (Fig. 7.23). We have

$$r_{max} = \lim_{m \to \infty} r, \quad K_{eff} = \left| \lim_{r \to 0} \frac{d(r/m)}{dr} \right|, \quad n_{eff} = \frac{\lim\limits_{r \to 0} (r/m)}{K_{eff}} \tag{7.43}$$

where r is the molar ratio of bound ligands to the number of base pairs of DNA and m is the concentration of free ligand in solution.

Statistical mechanics yields the following values:

$$r_{max} = \frac{n}{1+\psi/2}, \quad K_{eff} = K(1+\psi), \quad n_{eff} = \frac{n}{1+\psi} \tag{7.44}$$

$$n = x^{q_{AT}}(1-x)^{q_{GC}} \tag{7.45}$$

Here x is the fraction of A-T pairs in the polymer; K is the true association constant. The parameter ψ takes account of the decrease

in the number of possible positions on the polymer for a second ligand molecule due to the presence of the first molecule. The binding of successive ligands does not occur independently. The values of the parameter ψ are given in Fig. 7.24: (a) $l_{AT} = 0$; $l_{GC} = 0$; $\psi = 2(\alpha - 1)$; (b) $l_{AT} = 1$; $l_{GC} = 0$; $\psi = 2(\alpha - 1)x$; (c) $l_{AT} = 1$; $l_{GC} = 1$; $\psi = 2(\alpha - 2)x(1 - x)$; (d) $l_{AT} = 2$; $l_{GC} = 0$; $\psi = 2x + 2(\alpha - 2)x^2$; . . .; (e) $l_{AT} = \alpha$; $l_{GC} = 0$; $\psi = 2\sum_{i=1}^{\alpha-1} x^i$; . . . .

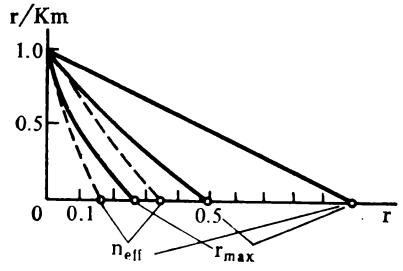


Fig. 7.23. The dependence of r/Km on r.

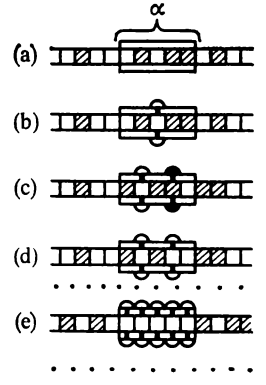


Fig. 7.24. The values of the parameter ψ.

Thus, for example, in the case of a homopolymer when binding is independent of the AT/GC ratio, we obtain:

$$K_{eff} = K(2\alpha - 1), \quad n_{eff} = \frac{1}{2\alpha - 1}, \quad r_{max} = \frac{1}{\alpha} \tag{7.46}$$

The difference between n_{eff} and r_{max} is quantitatively accounted for on the basis of the model adopted. The length of the ligand α can be found from experimental values of n_{eff} and r_{max}. From the isotherms of binding of ethidium bromide, spermine, actinomycin and netropsin by the DNA double helix the following values of α were obtained, respectively: 2, 4, 5, and 5 base pairs. These values are close to those expected on the basis of the molecular size of ligands.

Small molecules of the type of acridine dyes either slip in between base pairs in the DNA double helix (the *intercalation* model proposed by Lerman) or line up in a narrow groove on the surface of the double helix (the model proposed by Gursky). Evidently, both complexes are encountered. Preference cannot yet be given to either of the two models.

7.7. Replication of DNA

On cell division the genetic material is doubled and there occurs the reduplication (replication) of the DNA double helix. In principle, three mechanisms of replication are possible: *conservative replication* with the original DNA double helix remaining intact and a new daughter double helix being formed; *semiconservative replication*, in which two daughter double helices contain each one new and one old chain; and, finally, the dispersion mechanism involving the distribution of the original material among the four chains of two daughter double-stranded molecules. Meselson and Stahl (1958) studied the replication of DNA during cell division of *E. coli* with the aid of labeled atoms and sedimentation in the CsCl density gradient. They produced a population of ^{15}N-labeled cells. These cells were transferred to a medium containing ^{14}N exclusively, in which cell division took place. From the original population and from that of first and second generation (daughter and granddaughter) molecules there were isolated DNA molecules and their density and radioactivity were determined. The parental DNA molecule labeled with ^{15}N had the largest density; the DNA of the first generation (daughter molecules) was found to be half-labeled and its density was equal to the arithmetic mean of the densities of the ^{15}N- and ^{14}N-labeled DNAs. Finally, the DNA of the granddaughter molecules was separated on sedimentation into two zones—one zone with DNA being identical with the daughter molecule (half-labeled) and the other zone with unlabeled DNA. This may be depicted schematically as follows:

$$^{15}N\text{-}^{15}N \rightarrow 2\,^{15}N\text{-}^{14}N \rightarrow 2\,^{15}N\text{-}^{14}N + 2\,^{14}N\text{-}^{14}N$$

These results are in agreement with the semiconservative mechanism. It has also been confirmed by other data.

Template replication of DNA was also observed *in vitro* by Kornberg in 1960. An incubation mixture was taken that contained all four nucleoside triphosphates (ATP, GTP, TTP, CTP), Mg^{2+} ions (6×10^{-3} M), a polymerizing enzyme isolated from cells of *E. coli*, and native DNA (from calf thymus) as a primer; the pH value was 7.5. In such a system DNA is replicated—there are produced molecules identical with the primer DNA molecules. The process does not take place in the absence of one of the components or when the primer DNA is preliminarily treated with deoxyribonuclease, an enzyme that destroys DNA. It was demonstrated, with the aid of ^{32}P-labeled nucleotides, that the synthesis of DNA is the polycondensation of monomers—when one nucleotide is incorporated into the chain, one molecule of pyrophosphate, $H_4P_2O_7$, is released. According to the law of mass action, the reaction is retarded upon addition of an excess of pyrophosphate. If its amount exceeds by 100 times the total amount of the triphosphates, the reaction rate diminishes

by two times. The newly formed DNA has the biological properties of the parental DNA. For example, synthesis with DNA from phage ϕ X174 as the primer produces DNA which has an infective action—phage particles are synthesized. Using the polymerizing enzyme DNA-polymerase, one can produce synthetic polydeoxyribonucleotides capable of forming double helices in the presence of complementarity:

—A—T—A—T—A—T—A—T—
—T—A—T—A—T—A—T—A—

and

—G—G—G—G—G—G—G—G—
—C—C—C—C—C—C—C—C—

The free-energy change during the formation of two double helices from one parental double-stranded DNA and nucleoside triphosphates may be represented in the following form:

$$\Delta G = N\,(2E_1 + \overline{E}_2) - T\Delta S \tag{7.47}$$

where N is the number of nucleotides in the chain; E_1 is the energy released when one nucleotide is incorporated into the chain; $\overline{E}_2$ is the average value of the energy evolved as a result of interactions that stabilize a Watson-Crick pair. The energy $\overline{E}_2$ depends on the relative content x of G-C pairs:

$$\overline{E}_2 = xE_{GC} + (1-x)\,E_{AT} \tag{7.48}$$

The entropy change ΔS is determined by the replacement of a nucleoside triphosphate (NTP) by pyrophosphate (PP) in solution and by the loss of the rotational and translational degrees of freedom by the nucleotide incorporated into the chain. The latter factor is written in the form k ln q. On the whole

$$\Delta S = -2N\,\mathrm{k}\ln\left(q\frac{\overline{c}_{NTP}}{c_{PP}}\right) \tag{7.49}$$

where c_{PP} is the concentration of pyrophosphate in solution; $\overline{c}_{NTP}$ is the average concentration of the nucleoside triphosphate equal to $(c_A^{(1-x)}c_T^{(1-x)}\,c_G^{x}\,\,c_C^{x})^{1/2}$; c_A, c_T, c_G, and c_C are the concentrations of the respective nucleoside triphosphates. If they are equal, then $\overline{c}_{NTP} = 1/4c_{NTP} = c_A = c_T = c_G = c_C$. The replication condition consists in the liberation of free energy:

$$2E_1 + \overline{E}_2 + 2\mathrm{k}T\ln\left(q\frac{\overline{c}_{NTP}}{c_{PP}}\right) \geqslant 0 \tag{7.50}$$

The quantities E_1 and $\overline{E}_2$ are equal to about 16 and 20 kJ/mole, respectively. The quantity $q = Z_1'Z_2/Z_1Z_2'$, where Z_1 and Z_1' are the

partition functions for NTP in solution and for NTP bound to the template by hydrogen bonds; Z_2 and Z_2' are the partition functions for PP in solution and for PP present in NTP and bound to the template. Estimation gives $q \sim 10^{-2}$.

At 300 K the replication condition (7.50) is fulfilled until the ratio c_{PP}/c_{NTP} exceeds 400. In cells and in systems for synthesis the ratio is much lower. The condition (7.50) is valid if the amount of NTP is sufficient. Replication may be impeded by kinetic rather than thermodynamic factors.

An analogous treatment of denaturation (Sec. 7.4) leads to the following value of the difference between the free energies of the separated chains and the double helix:

$$\Delta G' = N\bar{E}_2 - 2NkT \ln q' \tag{7.51}$$

where q' characterizes the change in intrinsic entropy of the chain per one nucleotide. One can equate q' to the number of conformations of a unit of a single-stranded chain. The unit contains five single bonds, the rotation about which gives rise to two or three conformations. The order of magnitude of q' is 10-100. From Eq. (7.51) it follows that the temperature of the helix-coil transitions is equal to

$$T_m = \frac{\bar{E}_2}{2k \ln q'} = \frac{E_{AT}}{2k \ln q'} + x \frac{E_{GC} - E_{AT}}{2k \ln q'} \tag{7.52}$$

The experimental dependence of T_m (in K) on x

$$T_m = 342 + 41x$$

is obtained at $E_{AT} = 19.7$ and $E_{GC} = 22.3$ kJ/mole and $q' = 32$. These crude estimates are suitable only for the assessment of the order of magnitude of ΔH and ΔS.

Replication obviously requires the unwinding of the double helix. Replicative synthesis proceeds with the obligatory participation of *DNA-polymerase*, which is evidently displaced along the double helix, unwinding it and synthesizing new chains (the zip-fastener model). Bacteriophage ϕX174 contains a single-stranded rather than a double-stranded DNA. It has been found, however, that the multiplication of this phage proceeds through a replicative step, in which DNA becomes double-stranded. In this case too, it is exactly the double helix that undergoes replication. This also refers to viral RNA, the replication of which produces an intermediate double-helical structure. The single-stranded DNA of phage ϕX174 is circular, just like the double-helical DNA of other phages and bacteria. In the Kornberg experiments (page 268) this DNA served as a primer. However, only a circular DNA is biologically active, i.e., exerts an infective action. Kornberg made use of another enzyme, *ligase*, for closing a new chain. The new chains were built up not with thymine

but with bromouracil. In order to isolate synthetic DNA, the preparation was briefly treated with DNAase so that one break in the circle occurred in one half of the molecule. As a result, a mixture was obtained which contained, in equal amounts, template circles, synthetic circles, template and synthetic open DNAs, and also double circles without breaks. The sedimentation method in the density gradient produced synthetic circles containing bromouracil. These rings proved to be capable of infecting cells of *E. coli*. It had been established earlier that the replacement of even one of the 5500 nucleotides of DNA from phage ϕX174 led to the loss of infective action. Hence, the Kornberg synthesis is absolutely accurate.

The unwinding of the DNA double helix on replication (or on transcription in the synthesis of mRNA, see Sec. 8.3) must presumably be different from that involved in denaturation. The gain in free energy during the formation of new chains is responsible for a high angular momentum. The polymerase, which is displaced along the double helix, creates local conditions that disturb the balance of interchain interactions. It is supposed to hinder the hydrophobic interactions of the bases.

An elementary model for treatment of the kinetics of DNA replication is a one-dimensional template on which nucleoside triphosphates are adsorbed, their practically irreversible polycondensation taking place from one end. Crude estimates show that the diffusion of NTP from solution to the template proceeds faster than the processes occurring on the template, and the rate of synthesis is limited by the enzymatic process. Calculations indicate, in accordance with experiment, that the synthesis must proceed without an induction period.

We shall now take up a theoretical treatment based on the Maniloff model (1969). We have a template consisting of N cells. The growth centre (polymerase) is displaced from the first to the Nth cell. The synthesis is irreversible, i.e., the transfer of the polymerase from the $(i-1)$st to the ith cell takes place far from equilibrium. The rate constant of growth k_r for all cells is the same, the rate constant for initiation k_0 is much lower than k_r. The removal of the polymerase from the Nth cell occurs rapidly and does not limit the process. The task boils down to the consideration of irreversible random walk in a one-dimensional system. The kinetic equations are as follows:

$$\begin{aligned}
\dot{w}_1 &= k_0 - k_r w_1 \\
&\cdots\cdots\cdots \\
\dot{w}_i &= k_r w_{i-1} - k_r w_i \\
&\cdots\cdots\cdots \\
\dot{w}_N &= k_r w_{N-1}
\end{aligned} \tag{7.53}$$

where $i = 2, 3, \ldots, N-1$; w_i is the probability that the enzyme is in the ith cell. The number-average degree of polymerization is given by

$$\bar{n}(t) = \sum_{i=1}^{N} i w_i(t) \left[\sum_{i=1}^{n} w_i(t)\right]^{-1} \tag{7.54}$$

At the initial conditions $w_1(0) = \ldots = w_N(0) = 0$. The solution of the simultaneous equations (7.53) with the aid of the Laplace-Carson transformation has the form

$$w_i(t) = \frac{k_0}{k_r} \sum_{j=i}^{\infty} f(j), \quad i = 1, 2, \ldots, N-1$$

$$w_N(t) = k_0 t \sum_{j=N-2}^{\infty} f(j) - (N-1)\frac{k_0}{k_r} \sum_{j=N-1}^{\infty} f(j) \tag{7.55}$$

$$f(j) = (k_r t)^j \frac{\exp(-k_r t)}{j!}$$

We obtain

$$2k_r\bar{n}(t) = 2k_r t \sum_{j=0}^{N-3} f(j) + (k_r t)^2 \sum_{j=0}^{N-4} f(j) + 2Nk_r t \sum_{j=2}^{\infty} f(j) - N(N-1)\sum_{j=1}^{\infty} f(j) \tag{7.56}$$

At $k_r t \ll 1$ the probability $w_i \to 0$, $w_N \to 0$, $\bar{n} \to 1$, i.e., no synthesis takes place. At $k_r t \gg 1$ we find $w_i \to k_0/k_r$, $w_N \to k_0 t$, $\bar{n} \to N$, i.e., the synthesis goes to completion. The rate constant $k_r = pk_t$, where p is the probability that the cell that is in front of the growth centre contains the adsorbed monomer, and k_t is the rate constant of its insertion into the growing chain. Since the process is not limited by diffusion, p has an equilibrium value and is expressed by the Langmuir isotherm (cf. page 228):

$$p = \frac{a}{K+a} \tag{7.57}$$

where $K = \exp(E/\mathrm{k}T)$; E is the sorption energy; $a = \exp(\mu/\mathrm{k}T)$ is the absolute activity of NTP in solution; μ is the chemical potential of NTP. Since $\mu = \mu^0 + \mathrm{k}T \ln(\gamma c)$, $a = Ac$, $A = \gamma \exp(\mu^0/\mathrm{k}T)$ where γ is the activity coefficient, c is the concentration of NTP, and μ^0 is the standard chemical potential. At $p \ll 1$ the value of $a/K \ll 1$ and

$$p = a/K = Ac \exp(-E/\mathrm{k}T) \tag{7.58}$$

$$k_r = k_t Ac \exp(-E/\mathrm{k}T) \tag{7.59}$$

It is necessary that c be sharply increased for $\bar{n}$ to be increased from a small to a large value. Such an increase is not likely to occur on cell division. On the other hand, the dependence of $\bar{n}$ on E is found to be a sigmoidal curve and a slight change in E is sufficient for complete synthesis, which can be determined by the cooperativity of the process—by the interaction of neighbouring nucleotides on the template.

The above treatment is equivalent to a consideration based on the theory of absolute reaction rates (see page 191). The rate constant for the reaction of the i-mer that contains two nucleotides in the chain, may be written in the form

$$k = \frac{kT}{h} p \tag{7.60}$$

The experimental values for the rate of DNA synthesis are 1.8×10^3 monomers per second for *E. coli* and 2×10^2 monomers per second for mouse lymphoma cells. From this it follows that $p = 2.8 \times 10^{-10}$ for *E. coli* and $p = 3.1 \times 10^{-11}$ for mouse lymphoma. The evaluation of the free energy of activation for the adsorption of NTP but not for the insertion of a nucleotide into the chain, gives $G^{\ddagger} = 54.5$ kJ/mole for *E. coli* and 34.9 kJ/mole for mice. These values are found through the equilibrium constant for the activated complex of the n-mer with two nucleotides:

$$G^{\ddagger} = -RT \ln K^{\ddagger} = -RT \ln \frac{p}{\bar{n}c^2} \tag{7.61}$$

The theory described above is, of course, a simplified one, because it does not take into account either the template heterogeneity or the cooperativity of the process. Nonetheless, it gives a reasonable semiquantitative description of the template synthesis.

The enzymatic synthesis of polynucleotides can proceed without a template. In this case, the polymer is produced after a latent period, the duration of which decreases with increasing enzyme concentration. An example is the synthesis of poly-AT. If we introduce an oligomer as a primer, then the lag-period will decrease with its length. Even short oligomers may serve as templates for a growing polymer which slips off the template. The corresponding kinetic theory allows us to obtain the dependence of the time required for the synthesis of a polymer on its length and on the rate of separation of the chain from the template on which the polymer is growing. This rate falls off sharply when the length of the oligomer exceeds 4 nucleotides. Evidently, this is the critical size of the template at which a new chain can form a double helix with the template.

A new DNA is synthesized *in vitro* always in the 5′ → 3′ direction. At the same time, there are data which indicate that *in vivo* the two

chains are replicated—one in the 5′ → 3′ direction and the other in the 3′ → 5′ direction. Okazaki (1968) ascribed this to the discontinuous mechanism of replicative synthesis. Short sections of the new DNA are synthesized on one or both chains at the replicating point (replication fork), invariably in the 5′ → 3′ direction (Fig. 7.25). Then these sections are joined, forming new chains. If the model is correct, the last replicated sections of one or both daughter chains

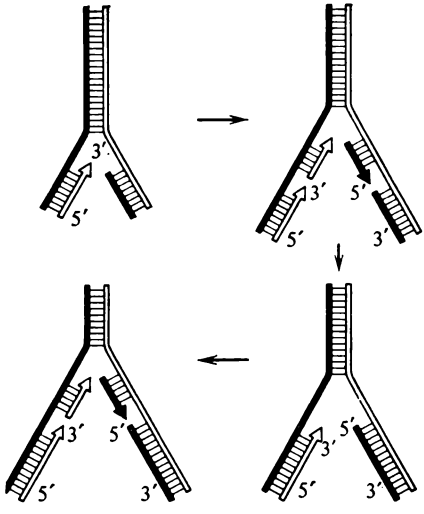


Fig. 7.25. Schematic representation of DNA replication according to Okazaki.

can be separated. The temporary selective inhibition of the enzyme that catalyzes the accumulation of phosphodiester linkages between the DNA chains must lead to the accumulation of new short chains. These predictions are verified by experiments which show that tritium-labeled thymidine is inserted into short chains, which only after a while unite into long ones.

The precise mechanism of DNA replication *in vivo* cannot as yet be considered to have been established. The replication *in vivo* starts at a definite stage of cell growth with the participation of factors that are evidently connected with the cell membrane (*the replicon model*).

Chapter 8 Physics of Protein Biosynthesis

8.1. The Problem of the Genetic Code

The gene is a part of the DNA macromolecule responsible for the synthesis of one protein chain. DNA contains information on the primary structure of the protein—the genetic information. It is contained in the primary structure of DNA, i.e., in the sequence of nucleotides. In viruses of the RNA type the genetic program is coded in RNA molecules.

What is the correspondence between the nucleotide sequence in DNA and RNA and the amino-acid sequence in the protein chain? This is what constitutes the problem of the genetic code.

This is a physical problem. As a matter of fact, there are four aspects to this problem: first, the correspondence between the information content of DNA and that of the protein; second, the quantitative relationships between nucleotides and amino acids which are determined in the long run by the interactions of molecules in the template synthesis of a protein—molecular recognition; third, the physical meaning of the genetic code; and, fourth, the evolutionary origin of the genetic code.

The problem of the genetic code was originally formulated by a physicist (Gamow, 1953) and a number of physicists took part in its elucidation. The code was deciphered, however, not by physical but by biological and chemical methods.

The protein text is written in the 20-letter language and the DNA (or RNA) text is written in the 4-letter language. From simple considerations it follows that the code ratio, i.e., the number of nucleotides that code for one amino-acid residue, cannot be less than 3. Indeed, the number of combinations of four nucleotides in groups of 2 is $4^2 = 16 < 20$ which is clearly not enough; if the nucleotides are considered in groups of three, there are 4^3 or sixty-four different combinations—an apparent excess.

Gamow assumed that the protein chain is built up directly on the DNA double helix, each amino acid being accommodated in the groove between the four nucleotides. This groove has a rhombic form. Two nucleotides belong to one chain and the other two to the other

chain. One of the nucleotides of the first chain forms a Watson-Crick pair with a nucleotide of the second chain. The Gamow "diamond code" provides just 20 "letters". Each "letter ", i.e., each rhomb, consists of four nucleotides. The number of combinations with the nucleotides considered in groups of four is $4^4 = 256$. But there are limitations, since the minor diagonal of the rhomb joins A with T or G with C. If the right-handed and left-handed forms of the rhombs are considered to be identical, for example,

$$\begin{matrix} & \mathrm{G} & & & \mathrm{G} & \\ \mathrm{A} & \cdots & \mathrm{T} & \equiv & \mathrm{T} & \cdots & \mathrm{A} \\ & \mathrm{A} & & & \mathrm{A} & \end{matrix}$$

we shall then have only 20 different rhombs.

The diamond code is overlapping. Since each rhomb incorporates nucleotides from three successive pairs, the two nucleotides that are on one side of the rhomb are common to two neighbouring rhombs. A correlation thereby appears between two neighbouring amino-acid residues. A particular amino-acid residue may be followed not by any of the 20 amino acids but by only some of them. However, further investigations have shown that the overlapping code cannot be reconciled with experiment, since any amino-acid residue may follow any of them. The absence of a correlation makes impossible a theoretical, speculative deciphering of the code.

The impossibility of an overlapping code also follows from the absence of mutation replacements of two neighbouring amino-acid residues in the protein. The mutation of a nucleotide common to two neighbouring residues must lead to such a double replacement.

The code ratio was found experimentally as a result of the genetic investigation carried out by Crick and his coworkers (1961), who studied the *r*II region of the genome of phage T4, which is multiplied in culture media of *E. coli*. It has been found that the mutations in this region caused by acridine dyes consist of the *deletion* of nucleotides and their insertion. The wild-type *w* is multiplied on the strains B and K12 of *E. coli*. The *r* mutants are multiplied only on B strains, forming plaques with sharp margins or halos. Some of the mutants of this type are capable of spontaneously reverting to the wild-type *w*. A genetic analysis has shown that such revertants result not from the reverse mutation $r \rightarrow w$ but from the appearance of the second suppressor mutation $w \rightarrow r$ near the first. Each of the two mutations leads to the loss of the ability to synthesize a corresponding protein, but the combination of two mutations in a single gene restores this ability. The number of *r* mutants that have been studied is about 80, including their double and triple combinations—suppressor suppressors and suppressor suppressor suppressors. All the suppressors were found to belong to two classes: + (addition or insertion of a nucleotide) and — (deletion). If the original *r* mutation is an insertion, its suppressor is then a deletion and vice versa. The wild phenotype gives

the following combinations: + —, — +, + + +, — — —, but not + +, — —, + + + +, — — — —.

These facts can be explained if we assume that the genetic code is a triplet code, nonoverlapping and readable in succession, starting from a certain fixed nucleotide. Let us depict the DNA chain by the sequence of the letters ABCABC... (Fig. 8.1*a*). Any of the four nucleotides, A, T, G, or C, may be in the position of each of the letters.

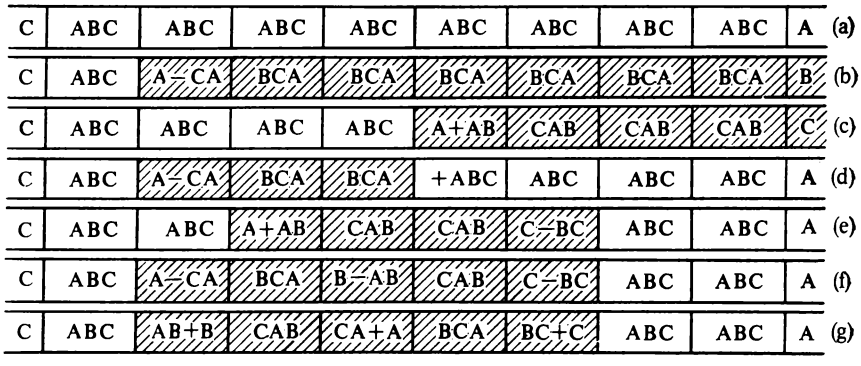

C ABC ABC ABC ABC ABC ABC ABC ABC A (a)
C ABC A—CA BCA BCA BCA BCA BCA BCA B (b)
C ABC ABC ABC ABC A+AB CAB CAB CAB C (c)
C ABC A—CA BCA BCA +ABC ABC ABC ABC A (d)
C ABC ABC A+AB CAB CAB C—BC ABC ABC A (e)
C ABC A—CA BCA B—AB CAB C—BC ABC ABC A (f)
C ABC AB+B CAB CA+A BCA BC+C ABC ABC A (g)

Fig. 8.1. DNA transcripts:
(*a*) wild-type; (*b*) mutant (deletion); (*c*) mutant (insertion); (*d*) and (*e*) double mutants of the + — type; (*f*) triple mutant — — —; (*g*) triple mutant + + +. The distorted portions are hatched.

The reading of the code, starting with a certain letter, is equivalent to superimposing a reading frame on this sequence. If one of the letters is deleted (—), or, on the contrary, a new one is inserted (+), then the entire sequence, beginning from the site of mutation, is distorted, i.e., a normal protein of the wild-type cannot be synthesized (Fig. 8.1*b* and *c*). If a new suppressor mutation has appeared (—+ or +—), then the sequence is distorted only in the region between the two mutations (Fig. 8.1*d* and *e*). If the distorted region is not too long, the protein being synthesized may retain its function and a reverse mutation or reversion will occur. Figure 8.1*f* and *g* illustrates reversions with three mutations of the same sign (either all deletions or all additions): — — — and + + +. It is also clear why double, quadruple, and quintuple single-type mutations do not produce reversions.

These results give a *code ratio* which is a multiple of 3. It is natural to assume that it is just equal to 3. It was later confirmed by a direct study of the primary structure of the synthesized protein. *Frameshift mutations* are also observed for the gene of phage T4

which produces lysozyme. It was found that shifts in the reading frame virtually distort the protein text. The double reverted mutant of lysozyme differs from the wild type in the sequence of five amino-acid residues:

The wild *w*-type . . . . . Thr Lys Ser Pro Ser Leu Asn Ala . . .
Revertant (pseudo-*w*-type) Thr Lys Val His His Leu Met Ala . . .

These genetic studies led to the solution of a number of physical problems. The code ratio has been established, the collinearity of the code proved, and it has been proved that the code is readout, beginning with a definite nucleotide, and has no commas, i.e., material units that separate the *codons*. Codons are triplets of successive nucleotides that code for amino acids. The total number of different codons is $4^3 = 64$.

In this connection there arose the question of the *degeneracy of the genetic code*. Since the total number of codons (64) by far exceeds the number of amino acids (20), it is necessary to find out how many different codons code for each amino acid. The term "degeneracy" is used here in the physical but not in the biological sense.

The investigation of frameshift mutations points directly to the degeneracy of the code. Let us return to the lysozyme fragments depicted above. If the code is not degenerate, then the two codons for His must be identical in the revertant. We designate them as ABCABC. If the shift is caused by addition of a nucleotide from the left and the codon for Pro is BCA, the relationship between the wild type and the revertant may be pictured thus:

... Pro Ser ...
... A BCA BCA ...
... ABC ABC ...
... His His ...

But it follows from this that the codon for Ser is also BCA. The contradiction is eliminated only if there are different codons for His.

The degeneracy of the code is also proved by the fact that with considerable variations in the DNA composition of different organisms (the fraction of G + C varies from 25 to 75 percent) the average composition of proteins is changed but slightly. At the same time, a number of facts point to the universality of the code.

The code degeneracy is also evidenced by the study of spontaneous mutations as well as those induced by chemical agents. Nitrous acid causes the replacement of the amino group by a hydroxyl group:

$$-NH_2 + HNO_2 \rightarrow -OH + N_2 + H_2O$$

In this way, the nitrite converts C to U, G to xanthine and A to hypoxanthine. After replication, xanthine and hypoxanthine are

replaced by G. Hence, the replacements C → U and A → G take place. Sixty-four triplet codons may be represented by eight octets (Wittmann). Figure 8.2 shows these octets; the arrows indicate the nitrite point mutations. The substitutions observed in the protein of tobacco mosaic virus (TMV) occur over the same octets. Wittmann has established that at least Ser or Ile (Fig. 8.3) must be coded for

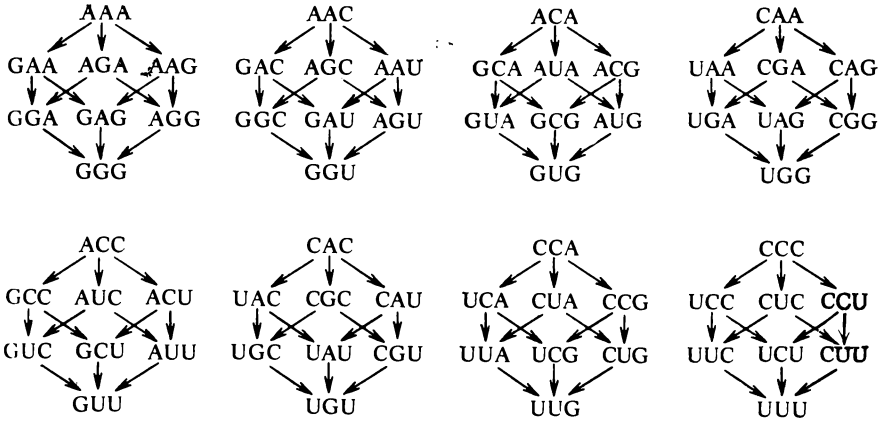


Fig. 8.2. The codon octets.

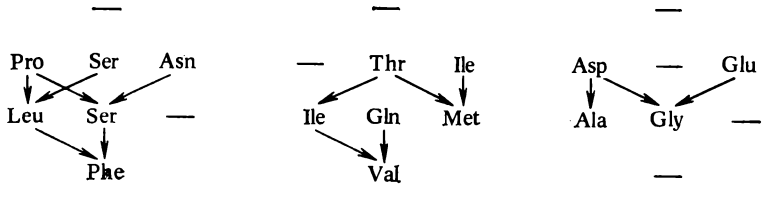


Fig. 8.3. The assignment of amino-acid residues to octets on the basis of nitrite mutations of tobacco mosaic virus (TMV).

by several triplets, i.e., the code is degenerate. These findings also demonstrate that most of the replacements in the protein occur as a result of the replacement of one nucleotide in the codon.

These propositions were later repeatedly confirmed. The genetic code is a triplet, nonoverlapping, commaless, and degenerate code. The code is universal for all organisms, from viruses to humans. [True, it has recently been found that in the synthesis of cytochrome oxidase in mitochondria the code is somewhat different from the

universal code (see page 298).] The complete deciphering of the code required direct biochemical experiments, which became possible only when the mechanism of protein biosynthesis was disclosed.

8.2. Protein Biosynthesis

The very existence of a fixed primary structure in the protein chain is evidence that the program of building up such a structure must have been encoded in the cell. The "text" cannot arise from the accidental encounters of the amino acids—like a typographical text, it must be "printed" on some "matrix". This was clear to Koltsov, long before the discovery of the role of nucleic acids. He maintained that the role of the "matrix" (template) responsible for protein synthesis is played by the protein itself. Today we know that the DNA and RNA molecules serve as templates. In order "to print a text", a genetic code is required. The template principle of protein biosynthesis is fundamental to molecular biology and molecular biophysics.

The information contained in genes, i.e., in DNA, is transferred to mRNA during *transcription.* Protein synthesis takes place exactly on mRNA. The so-called *central dogma* of molecular biology states that genetic information flows from DNA to RNA to proteins, a relationship symbolized as

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein} \qquad \text{(A)}$$

The genetic information can never be transmitted from the protein to the nucleic acids.

In 1970 Temin and Baltimore discovered *reverse transcription*, which is the transmission of the genetic information from RNA to DNA:

$$\text{RNA}' \rightarrow \text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein} \qquad \text{(B)}$$

The existence of the reverse transcription (see Sec. 8.3) in no way contradicts the central dogma. As a matter of fact, this is a law of nature rather than a dogma.

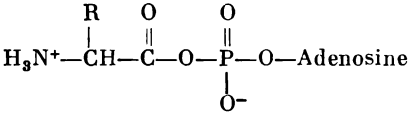
The template, or messenger, RNA transmits the genetic information from the *chromosomes*, where it is stored, to the *ribosomes*, on which the biosynthesis is accomplished. For protein biosynthesis to be accomplished, the starting material—the amino acids—is required and it is also necessary that the appropriate thermodynamic and kinetic conditions be fulfilled.

The amino acids are present in the cell in a free state. The direct polycondensation of the amino acids is an endergonic process—it is accompanied by an increase in the free energy by about 12 kJ/mole in the course of formation of a peptide bond. In the cell, the polycondensation of amino acids is coupled with the exergonic reaction of

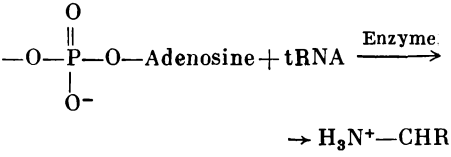
dephosphorylation of ATP (see page 43). An amino acid enters into the biosynthesis reaction in an activated form:

$$\text{Amino acid} + \text{ATP} \xrightarrow{\text{Enzyme}} \text{Aminoacyl adenylate} + \text{PP}$$

The symbol PP stands for pyrophosphate. Aminoacyl adenylate, in which the amino acid is activated, has the structure



The enzyme aminoacyl-tRNA synthetase also catalyzes the reaction leading to the formation of aminoacyl adenylate, and the transport of the thus activated amino acid to the terminus of the tRNA adaptor-molecule. The scheme of the second reaction is as follows:



The energy required for biosynthesis is stored in the chemical bond between the amino acid and tRNA.

For each amino acid there exists one or more specific tRNA's (see Sec. 8.4).

Protein biosynthesis occurs at ribosomes. The ribosomes are nucleoprotein particles composed of proteins and *ribosomal RNA* (rRNA). The ribosomes provide the appropriate interaction between mRNA and tRNA, which carry the amino acids, and the polycondensation of the amino acids to form a polypeptide chain. The ribosome plays the role of a catalyst, similar to that of an enzyme (see Sec. 8.5).

The ribosome attaches itself to the 5′-end of mRNA, which codes the N-terminus of the polypeptide chain. The ribosome moves along the mRNA chain, "reading the text" codon after codon in such a manner that the codons in the mRNA-ribosome complex bind tRNA's linked to the amino acids. Each of the tRNA's, carrying an amino acid, interacts in a complementary way with the codon of mRNA by its *anticodon*. The anticodon is a triplet that is complementary to the mRNA codon. The scheme of the process is shown in Fig. 8.4. On the ribosome there occurs the successive attachment of two aminoacyl-tRNA complexes to the nth and $(n + 1)$st codons of mRNA. A peptide bond is formed between the corresponding amino acids; the growing protein chain is transferred from the nth aminoacyl-tRNA complex to the $(n + 1)$st one; the nth tRNA splits off and the ribosome that carries the $(n + 1)$st tRNA with the growing chain

attached to it is moved by one codon. In the new position the ribosome interacts now with the $(n + 1)$st and $(n + 2)$nd codons of mRNA. Several ribosomes rather than one ribosome move along a single mRNA chain and each of the ribosomes has its own protein chain growing on it. The mRNA-ribosome system, which resembles a string of beads, is called a *polyribosome*, or a *polysome*. Thus,

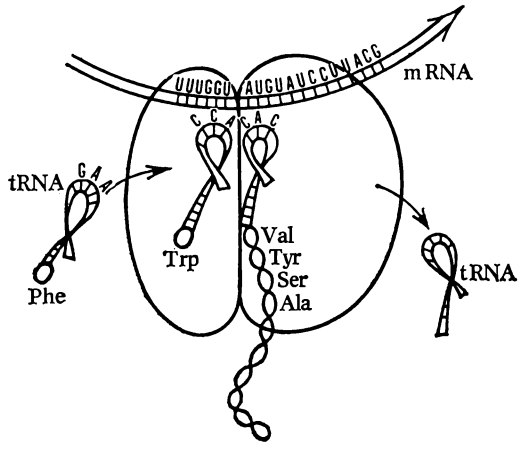


Fig. 8.4. Binding of amino-acyl-tRNA to mRNA on the ribosome.

a number of identical protein chains are synthesized on a single mRNA chain. The synthesis of a protein chain on a polysome is known as *translation*—the conversion of the nucleotide text to the amino-acid text, following the dictations of the genetic code, i.e., the codon-amino acid dictionary. The overall scheme of protein synthesis is presented in Fig. 8.5.

Thus, the protein chain is synthesized *de novo* from the amino acids and not from the previously existing polypeptide blocks. Then, the chain (or several chains) gives rise to a biologically functional protein molecule.

The above-described sequence of events in protein biosynthesis is complicated, but all its steps are characterized by a single principle that forms the basis of molecular biophysics. This principle is *molecular recognition*, which is accomplished via weak interactions (see page 62). The biosynthesis involves four types of nucleic acids: DNA, mRNA, rRNA, and tRNA. The pairs that recognize each other are as follows: the nucleotides of the complementary DNA chains; the nucleotides of DNA and mRNA; the codons on mRNA and the anticodons on tRNA. The recognition is effected also on interaction of nucleic acids with enzymes and ribosomes.

The precision of recognition of nitrogenous bases is demonstrated in Table 8.1. The complementary pairs A-U and G-C prove to be the most stable; for example, A-U is much more stable than A-A or U-U. The formation of noncomplementary pairs, however, is also possible, this possibility being one of the most important causes for the generation of mutations.

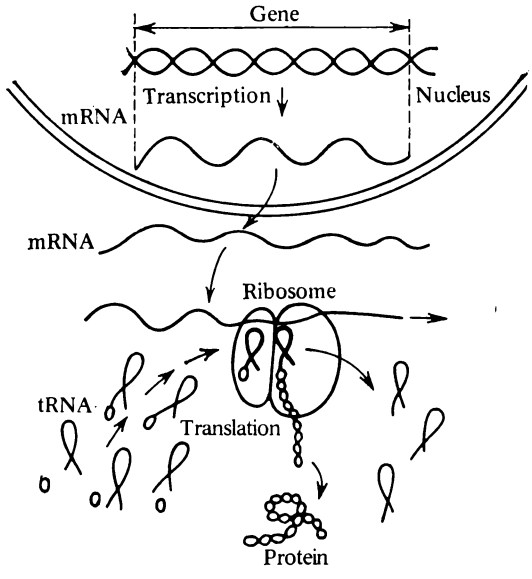


Fig. 8.5. The general scheme of protein synthesis.

The study of associations of oligonucleotides has shown that the pairing of triplets is most favourable for the biological function. Doublets possess too low values of K and quartets give, on the con-

Table 8.1. The Association Constants K (mole^{-1}) for Base Pairing in Nonpolar Solvents C_6H_6 and CCl_4 at 25°C

Base	U	A	C	G
U	15 *45*	*550*	*50*	$< 10^3$
A	150	8 *22*	*50*	$< 10^3$
C	< 28	< 28	28 *50*	$> 10^3$
G	$< 1.12 \times 10^3$	$< 1.2 \times 10^3$	3×10^4	1.2×10^3 ca. *10^3*

The *2'*-, *3'*-, and *5'*-O-substituted ribonucleosides have been studied; the italicized numbers refer to CCl_4 and the nonitalics to C_6H_6.

trary, too strong binding. The lifetime of a codon-anticodon pair must not exceed several milliseconds, since otherwise it will limit the rate of functioning of the biological system.

Thus, the complicated interplay of weak forces leads to the formation of strong chemical peptide bonds. The formation of chemical bonds and mechanical processes—the transfer of the enzyme *RNA-polymerase* along the DNA chain (see Sec. 8.3) and the action of the ribosome and its movement along the mRNA chain (see Sec. 8.5)—require the expenditure of energy. The energy sources for these processes are the macroergic nucleoside triphosphates ATP and GTP.

The steps of protein biosynthesis and their participants are considered in more detail below.

8.3. Transcription and Reverse Transcription

The transcription is a process by which the information contained in DNA is conveyed to RNA, i.e., the synthesis of RNA on a DNA template, this process being accomplished with the aid of RNA-polymerase. The information encoded only in one of the two strands that make up the DNA double helix is transcribed. The RNA-polymerase has a complex structure. This enzyme performs a number of functions: the recognition of the initial site of synthesis and the specific binding of DNA at this site; the initiation of RNA synthesis; the synthesis of RNA; the termination of the synthesized chain and the separation from the template.

At low ionic strengths RNA polymerase is a dimer with a molecular weight of about 10^6 and a sedimentation coefficient of 23S. At ionic strengths higher than 0.1 M, the enzyme breaks down reversibly into two 13S monomers. The 13S monomer consists of four subunits: β (molecular weight 150,000), β' (165,000), two α-subunits (each with a molecular weight of 40,000), and the σ-factor (80,000). The complex structure of RNA polymerase is associated with the large variety of its functions.

RNA polymerase is more active with a double-helical DNA rather than with a single-stranded DNA. The transcription *in vivo* occurs with a double-stranded DNA, though, as has been pointed out, only one chain is read. Accordingly, a local unwinding of the double helix takes place at the point where the RNA chain is growing. The application of the formaldehyde method, which allows the defects in the DNA double helix to be detected, has really assisted in revealing its unwinding on interaction with the enzyme. The rates of the synthesis of RNA *in vivo* and under optimal conditions *in vitro* are almost the same and range from 20 to 30 nucleotides per second.

Thus, the transcription involves the conformational change of DNA, its unwinding. RNA polymerase is located very sparsely on DNA and these changes involve only a small portion of the double

helix. One can, however, study the conformational properties of the complexes of RNA polymerase with small portions of DNA. Such complexes are produced on treatment of the DNA-RNA polymerase complex with the enzyme DNAase.

Florentiev and Ivanov have proposed a model for the action of RNA polymerase (1970). When the RNA chain is growing, the 3′-hydroxyl of the ribose is bound to the 5′-phosphate of the attacking nucleotide. It is presumed that when DNA is unwound and a hybrid DNA-RNA double helix is formed, the DNA template functions not in the ordinary *B*-form but in the *A*-form. With each step of the transcription process a new base is "extracted" from the DNA template, and the deoxyribonucleotide, whose complementary ribonucleotide has lost the pyrophosphate, moves back to DNA with the hydrogen bonds being re-formed. The axis of the hybrid double helix moves along the DNA duplex in a spiral manner, this being accompanied by the movement of the enzyme. Figure 8.6 is a schematic representation of a double-stranded DNA and the growing end of RNA during the action of RNA polymerase. The energetics of the transcription process is reduced to the coupling of the exergonic and endergonic steps. As a matter of fact, we are dealing here with the electronic-conformational interactions—the interaction of chemical and conformational degrees of freedom. The model shown is consistent with experimental data.

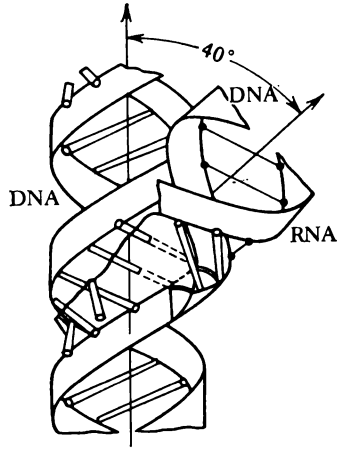


Fig. 8.6. The structure of the double-helical DNA template and of the growing end of RNA when acted on by RNA-polymerase.

The RNA-directed synthesis of DNA or reverse transcription, discovered as one of the steps of multiplication of RNA-containing oncogenic viruses (see page 280), is of great importance to molecular biology and molecular biophysics. Reverse transcription too invariably involves an enzyme called *reverse transcriptase* or *revertase*. The reverse transcriptase has three functions: the DNA polymerase function, which provides the synthesis of a complementary strand of DNA (cDNA) on the RNA template; the polymerase function, which ensures the synthesis of the second strand of DNA, *anti*-DNA, which forms a double helix with cDNA; and the ribonucleoside activ-

ity, which hydrolyzes RNA in the double RNA-DNA helix. The reverse transcriptase performs its functions with any RNA.

Reverse transcription makes possible the solution of important problems. By converting the amino-acid sequence of RNA into DNA, it is possible to disclose very accurately and rapidly the primary structure of DNA. The method proposed by Sanger in 1975 may be described by the following scheme:

RNA → cDNA → double-stranded DNA → recombinant DNA →
→ restriction enzyme fragments → analysis of fragments

Recombinant DNA's are produced by way of the multiplication of the resulting double-stranded DNA by using methods of *genetic engineering. Restriction*—the cleavage of the double helix at definite sites—is effected with the aid of special enzymes called collectively *restriction enzymes.* The fragments produced are analyzed by means of electrophoresis in a polyacrylamide gel. Today, it is possible to determine the sequences of many thousands of nucleotides in DNA, i.e., the structures of the genes.

Genetic engineering owes its origin to the reverse transcription. Complete double-stranded structural genes are synthesized and inserted into *plasmids* and phages, with the aid of which DNA molecules are multiplied. The transfer of genes is accomplished by means of DNA transcripts. Reverse transcription enables one to study the genomes of viruses, to decipher molecular defects that are responsible for the hereditary blood diseases, and to study molecular mechanisms that control the transformation of normal cells to tumour cells.

The structures of reverse transcriptases, large-size proteins, have not been worked out so far. The native reverse transcriptase of the oncornaviruses of birds has a sedimentation coefficient of 7.5S. Reverse transcriptases have a quaternary structure.

DNA polymerases, RNA polymerases, and reverse transcriptases invariably contain Zn^{2+} ions as cofactors. This is another indication of the importance of bioinorganic chemistry for biology (see page 53).

8.4. Transfer RNA

Transfer RNA's (tRNA's) constitute about 10 percent of the total amount of RNA in the cell. tRNA's were the starting point in the study of the structure of nucleic acids (Holley, Baev). At present, the primary structures of about 100 molecules of tRNA's of different origin are known. There are 20 families of tRNA's corresponding to the 20 amino acids.

The secondary structure of various tRNA's is, in principle, of the same type. Obviously, this must be so for tRNA molecules to perform their functions—the recognition of the mRNA codon on the ribosome

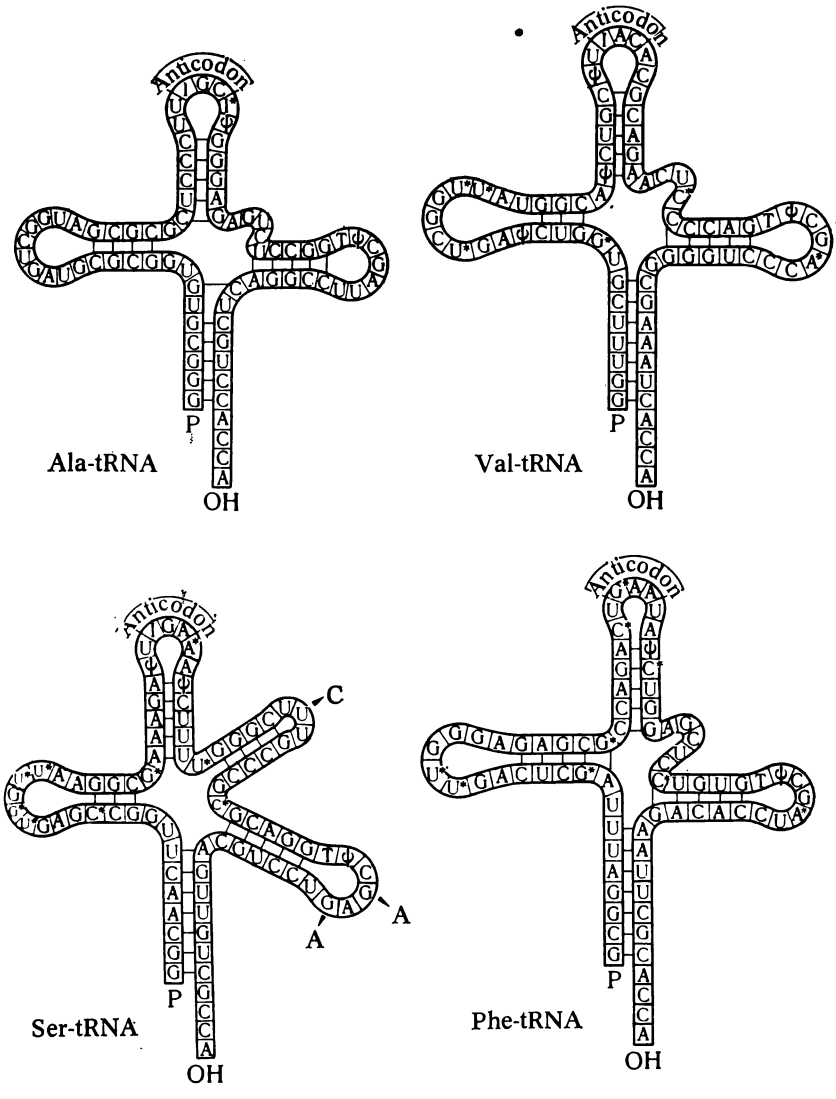


Fig. 8.7. The secondary structures of some tRNAs (ψ, A*, U*, G*, and I are minor nucleotides).

and the amino acid incorporation into the protein chain. Only the initiating methionine tRNA has a somewhat different structure.

The secondary structure of tRNA may be spectacularly depicted on the basis of the possibility of pairing of the complementary bases (A with U and G with C) and assuming that the nonpairing portions are loops (page 245). Then all known tRNA's may be depicted by the *clover-leaf model* with four or five double-helical portions (Fig. 8.7).

The four portions are constant, and the fifth, usually the smallest, varies. The clover-leaf model gives the projection of the tRNA molecule on the plane and reflects only its topology, i.e., the disposition of the double-helical sections.

The problem of working out the secondary structure of RNA from the nucleotide sequence was studied by Tumanian (1966). The algorithm includes two steps. First, there is constructed a triangular template of all possible pairs according to a simple rule: the element a_{ij} which is at the intersection of the ith column and the jth line is equal to 1 or 0 if the ith and jth nucleotides are complementary or noncomplementary, respectively. In order to take into account the difference in free energy between the G-C and A-U pairs, the corresponding weights are included instead of 1. The second step is the construction of "maximum hairpins", i.e., the elucidation of the secondary structure with a single double-helical region with a maximum negative free energy. What remains to be done is to run over a limited number of secondary structures with different numbers of maximum hairpins found.

Later (1971) Tinoco obtained more accurate values of free energies of the formation of various types of pairs and loops.

As regards the secondary structure of tRNA, it can be found, as has been shown by Eigen, with the aid of the game model. "The tRNA game or how to make clover leaves" consists in the following. Each player is given a random sequence of N units belonging to four classes: A, U, G, and C, and a tetrahedral die, each face of which corresponds to one of these letters. The players throw a die by turns and, filling a certain place in the sequence with the letter cast, each player strives to obtain a double-stranded structure with a maximum number of A-U and G-C pairs. The game is finished when one of the players declares that he has obtained the "complete" structure. The winner is the player who got by this moment the maximum number of points. The points are counted for pairs; the player can be given, for example, two points for a G-C pair and one point for an A-U pair. The pairs can be formed only if there is a continuous sequence of 2 G-C or G-C, 2 A-U or 4 A-U (the cooperativity rule). For each loop in the structure at least 5 positions must remain unpaired (cf. Fig. 8.7). The player throws a die for any position in the sequence declared by him in advance. At $N = 80$ there is always obtained a structure of the clover-leaf type with 3-4 leaves. The hairpin that has only one

loop, contains a maximum number of bases that can be paired. The clover-leaf model makes it possible, however, to test a much greater number of combinations, since separate leaves can be shifted independently and from the outset a much larger number of combinations are possible. Eigen remarks that nature has evidently been playing this game since time immemorial. A game of a similar type can perhaps be devised for solving the problem of the relationship between the primary and the three-dimensional structure of the protein (Sec. 4.6).

The three-dimensional structure of native tRNA is rather compact. Thermally denatured tRNA is capable of renaturation. Rich (1972) succeeded in deciphering the structure of phe-tRNA by means of X-ray structural analysis. This tRNA is built up of two helical segments linked at a right angle, like the letter Γ. The structure is shown in Fig. 8.8. The compact structure of tRNA is stabilized by horizontal and vertical interactions in the double-helical portions, which are packed tightly under the influence of hydrophobic and electrostatic forces. The Mg^{2+} ions required for maintaining the native structure compensate for the negative charges of the phosphate groups.

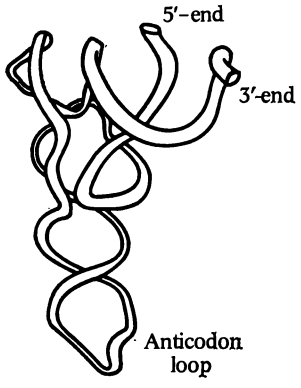


Fig. 8.8. The three-dimensional structure of phenylalanine-tRNA (after Rich).

The CD spectrum of tRNA has intensive positive maxima at 264 nm ($\Delta\varepsilon$ about 5.5) and at 188 nm ($\Delta\varepsilon$ about 18), negative maxima at 296 nm ($\Delta\varepsilon = -0.35$), 233 nm ($\Delta\varepsilon = -1.0$), and 210 nm ($\Delta\varepsilon$ about −5.0). These bands are sensitive to conformations.

The study of the migration of energy between the luminescent base present in the anticodon loop of the phe-tRNA molecule and the chromophores (acridine dyes) linked covalently to the acceptor end, has made it possible to estimate the distance between them in accordance with structural data.

8.5. Translation

Translation, which is the conversion of the polynucleotide transcript of DNA and mRNA into the amino-acid, protein text occurs in complexes of ribosomes with polyribonucleotide chains, i.e., in polysomes.

Ribosomes can only be observed with the aid of an electron microscope (Fig. 8.9) as spherical particles with a diameter of 15-25 nm. The characteristics of ribosomes are listed in Table 8.2.

The ribosomes of eukaryotes are characterized by a high value of $S^{0}_{20,w}$ (80S and not 70S) and by molecular weights of (4.1-4.7) $\times$ 10^6

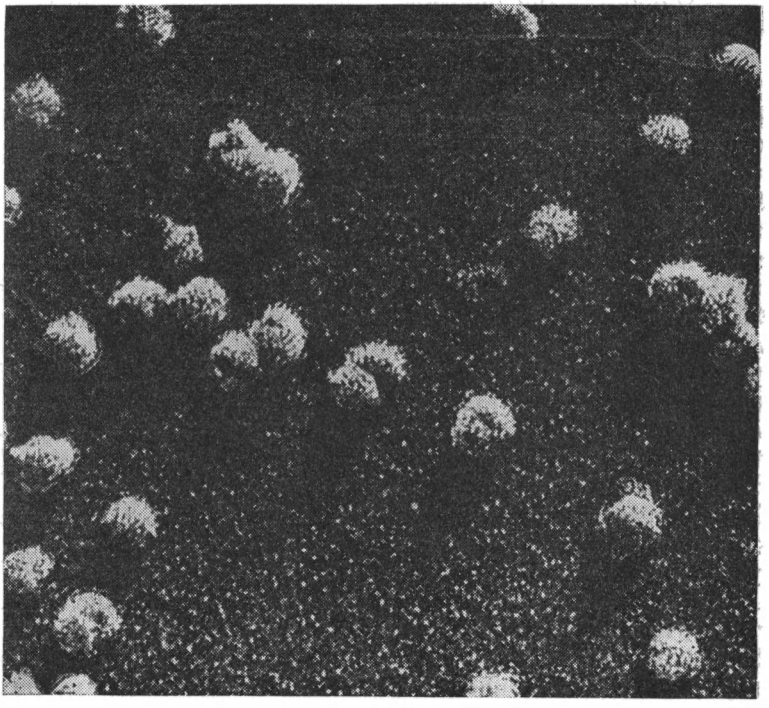


Fig. 8.9. An electron micrograph of ribosomes (after Vasiliev). Magnification 500,000×.

and by somewhat larger sizes (the dry volume is 5×10^3 nm^3). The 70S ribosomes consist of two subunits, 30S and 50S, into which they break down with a decrease in the concentration of divalent ions (Mg^{2+}) or with an increase in the concentration of univalent ions. The molecular weight of 50S particles is 1.8×10^6 and that of 30S particles is $(0.7\text{-}1.0) \times 10^6$. The shapes and sizes of 70S, 50S, and 30S particles are given in Fig. 8.10 (cf. Fig. 8.9).

Ribosomal RNA constitutes about 65 percent of the dry weight of ribosomes, the remaining 35 percent being proteins. These RNA's

Table 8.2. The Physical Properties of *E. coli* Ribosomes

1. Sedimentation coefficient $S^0_{20,w}$ (in Svedberg units)	69.1-70.5
2. Intrinsic viscosity [η], cm^3/g	6.1-6.8
3. Translational diffusion coefficient $D^0_{20,w} \times 10^7$	1.83
4. Specific volume, cm^3/g	0.64-0.60
5. Molecular mass	ca. 3×10^6
6. Size in dried state, nm	ca. $20 \times 17 \times 17$
7. Volume, nm^3	ca. 3×10^3
8. Size in aqueous medium, nm	ca. $30 \times 30 \times 20$
9. Volume, nm^3	ca $(3\text{-}10) \times 10^3$
10. Amount of bound water, g/g	0.9

are subdivided into 3 classes: 23-28S, molecular weight $\geqslant 1 \times 10^6$; 16-18S, molecular weight $< 1 \times 10^6$; and low-molecular-weight RNA's—5S, molecular mass about 40,000. Presumably, the protein molecules interact with the nonhelical portions of rRNA and the ribonucleotide-protein complex is coiled into the compact structure of the ribosomal subunit. The 70S ribosome contains about 65 polypeptide units with an average molecular mass of 65,000. The 30S particles contain 19-20 species of proteins, while the 50S particles contain more than 50 kinds of proteins.

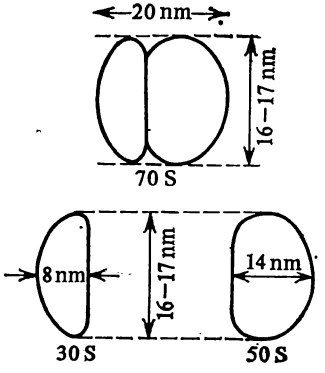
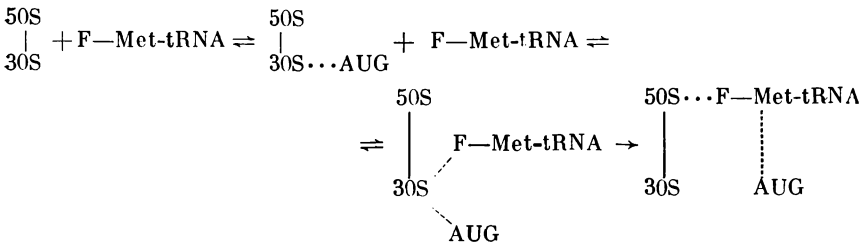


Fig. 8.10. Schematic representation of 70S, 50S and 30S particles.

The complete "disassembly" of ribosomes and their reconstitution from the proteins and rRNA has been accomplished by Nomura (1966-1969). The role of rRNA is specific but not absolute—the functional 30S particles are produced from the 16S-rRNA of one species of bacteria and ribosomal 30S proteins of another species. The number of proteins isolated was 19. The absence of any of these affects the function of the 30S particle. All the proteins function in a concerted manner, which is also necessary for *self-assembly*. The kinetics of self-assembly corresponds to a first-order reaction, which implies the presence of a certain slow intermediate step which limits the reconstitution process. The strong dependence of the rate of self-assembly on temperature shows that for this reconstitution to be affected a free energy of about 160 kJ/mole of ribosomes is needed. The self-assembly of ribosomes from all the components *in vitro* takes about 5 min.

For the ribosomes to function properly in protein biosynthesis, it is necessary that special proteins be present in them, namely *transfer factors*, which bind to the ribosome in the presence of divalent cations. These factors have been studied little. Divalent ions (Mg^{2+} in a concentration not less than 0.005 *M*) are also required for the interaction of the ribosomes with mRNA or with artificial polyribonucleotides. Each ribosome unites only with one polynucleotide chain. The binding occurs in the 30S subunit. In the triple ribosome-mRNA-aminoacyl-tRNA complex, the third component becomes bound by its anticodon to the mRNA codon. This recognition and pairing also occur in the 30S subunit.

Translation begins with *initiation*, i.e., the synthesis of the first peptide linkage. Two amynoacyl-tRNA complexes are simultaneously involved in the initiation. One of them is the first one. It has been found that the first complex formed is invariably N-formylmethionyl-tRNA (symbolized as FMet-tRNA), which has an enhanced affinity for the peptidyl-tRNA binding site of the ribosome on the 50S subunit. The initiating mRNA codon is AUG located at the 5′-end of the polynucleotide (see page 297). Admittedly, this codon is preliminarily bound by the aminoacyl-tRNA binding site of the 30S subunit. The overall scheme of the initiation process is as follows (F is formyl):



This is followed by the consecutive polycondensation of the amino acids. During the entire course of the translation process the growing polypeptide is kept bound to the ribosome. Attachment of each of the subsequent aminoacyls occurs at the C-end of the polypeptide chain. The transfer RNA that brings an aminoacyl remains to be bound to it. This aminoacyl arranges itself by way of the replacement of the tRNA by the aminoacyl-tRNA complex. We obtain the following scheme:

$$\text{Peptidyl } (n)\text{-tRNA}' + \text{aminoacyl-tRNA}'' \rightarrow \text{Peptidyl } (n+1)\text{-tRNA}'' + \text{tRNA}'$$

This process repeats itself many times on each ribosome. Spirin proposed a spectacular model for the operation of the ribosome (Fig. 8.11). It is assumed in this model that the aminoacyl-tRNA binding site and the peptidyl-tRNA binding site are localized at dif-

ferent subunits, at 30S and 50S, respectively (Fig. 8.11*b*). The peptidyl-transferase centre, which provides the transfer of the peptidyl, resides at the 50S particle. The ribosome is periodically opened and closed. Figure 8.11*c* shows a closed ribosome, with the aminoacyl end of the aminoacyl-tRNA complex being located very close to the esterified carboxyl of the peptidyl. The closing of the ribosome

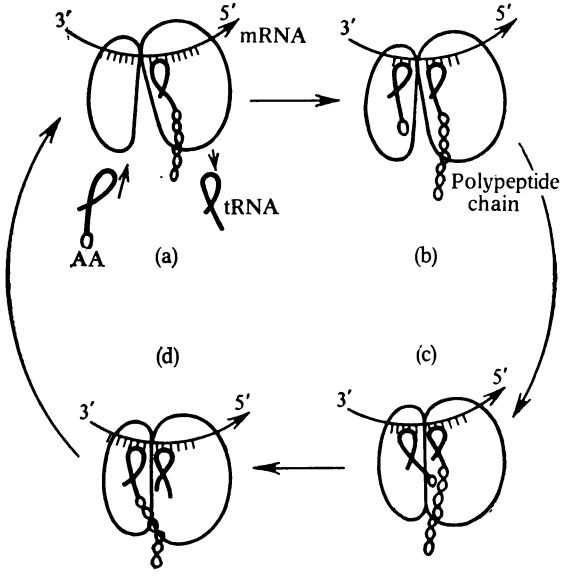


Fig. 8.11. Schematic of the operating cycle of the ribosome according to Spirin

subunits ($b \rightarrow c$) is accompanied by the formation of a peptide linkage. The carboxyl is transferred to the amino group of the aminoacyl-tRNA complex, and the deacylated tRNA remains bound to the 50S particle (Fig. 8.11*d*). Then translocation takes place—the tRNA is transferred from the peptidyl-tRNA from the 30S to the 50S particle, carrying with it the mRNA codon linked to it and displacing the deacylated tRNA from the peptidyl-tRNA binding site. The ribosome is again opened (Fig. 8.11*a*). Then a new aminoacyl-tRNA complex becomes attached to the new codon in the 30S particle and the cycle repeats. The periodic opening and closing of the ribosome is, according to Spirin, the driving mechanism that provides spatial displacements of tRNA and mRNA during the translation process. The periodic changes in the quaternary structure of the ribosome are transformed into the translational motion of the mRNA chain.

This is a mechanochemical process—mechanical work is done at the expense of the energy released upon binding by weak forces and at the expense of the chemical energy of GTP. The closing of the ribosome is induced by the incoming aminoacyl-tRNA being attached to the ribosome, while its opening requires the energy of GTP. Gavrilova and Spirin have shown that the ribosome can also operate *in vitro* without GTP. Probably, the ability to display such a *nonenzymatic translation* is embedded in the structural organization of the ribosome.

No quantitative physical theory of all these events has been worked out as yet. It has been shown experimentally that the translation depends on the conformation of mRNA and that the active ribosome complex really undergoes conformational changes during translation.

As has already been said, one mRNA chain combines with a number of ribosomes, forming a polysome. The size of polysomes depends on the length of the mRNA chains. There is one ribosome per every 80-90 nucleotides in the chain. In the synthesis of proteins containing about 150 amino acids the polysome has 4-6 ribosomes, while in the synthesis of longer protein chains the polysome can contain 12-20 and more ribosomes. One mRNA chain thus provides the synthesis of a number of protein chains. The template RNA undergoes then degradation under the action of the enzyme *exonuclease*, which has been studied little. The ribosome, being attached to the 5′-end of mRNA, protects it from degradation. When the ribosome is moving, the 5′-terminal portion either breaks down or binds a new ribosome. The degraded end can no longer bind a ribosome, but the earlier attached ribosomes continue to move and synthesize new protein chains. The time required for the synthesis of a protein chain containing 400 amino-acid residues is about 30 s. The time τ during which the ribosome travels the average distance δ between two ribosomes in the polysome is about 3 s, which corresponds to a linear velocity of the order of 10^{-6} cm/s.

A mathematical model that describes these processes allows one to calculate the distribution of the sizes of mRNA in agreement with experiment.

On the whole the protein biosynthesis is determined by the rates of transcription and translation. Presumably, these rates are commensurate, since with mRNA labeled with radioactive atoms there is observed a correlation between the lengths of labeled mRNA's and polysomes and there are no free labeled mRNA's present.

8.6. Deciphering of the Genetic Code and Its Meaning

A direct way towards the solution of the problem was suggested by Nirenberg and Matthaei (1961). Upon introduction of synthetic polynucleotides into a cell-free system the amino acids are incorporated into the polypeptide chain. The cell-free system contained ri-

bosomes, a set of tRNA's, ATP, all the necessary enzymes, but it did not contain DNA and mRNA. The system was prepared from degraded cells of *E. coli*. The ribosomal and supernatant fractions were isolated by centrifugation. The ribosomes were washed off and the supernatant liquid containing enzymes and tRNA was subjected to dialysis against a special buffer. To the mixture of these two purified fractions was added an ATP-generating system—a source of energy. Synthetic polynucleotides were injected into the cell-free system, and the incorporation of ^{14}C-labeled amino acids into the fraction insoluble in trichloroacetic acid, i.e., polypeptides, was studied. It turned out to be possible to "deceive" the biological system —the ribosomes interacted with synthetic polynucleotides instead of with a natural mRNA. It was found that the poly-U chain stimulates the polycondensation of phenylalanine, and poly-C performs the same function with respect to proline.

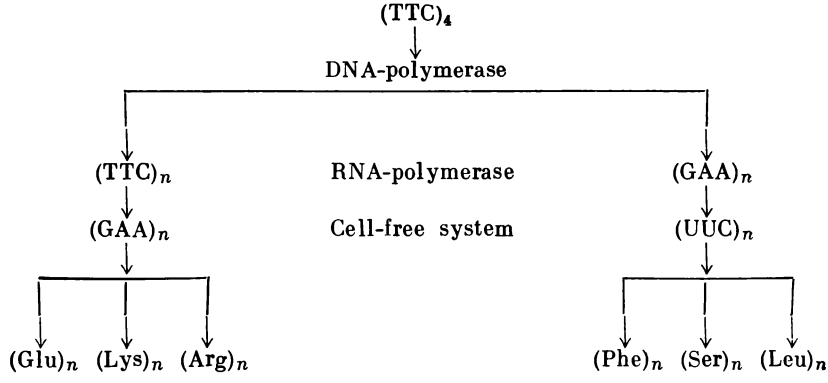
Thus, the first codons were found. If the code is a triplet, then the codon UUU corresponds to Phe, and the codon CCC to Pro. Then, there was studied the action of copolymers of known composition but with a random sequence of nucleotides. The frequency of appearance of triplets of definite composition is known for such copolymers. The copolymer AU in the proportion 1 : 5 stimulates the incorporation into the polypeptide chain of Phe, Leu, Ile and also, in smaller amounts, of Asn and Lys. If 100 is the fraction of the triplet 3U, then the fraction of 1A2U is 20 (3U/1A2U = 5), the fraction of 2A1U is 4 (3U/2A1U = 25), and that of 3A is 0.8 (3U/3A = 125). Comparing these values with the amount of incorporation (for Phe it is assumed to be equal to 100), one can find the composition of the codons for the amino-acid residues indicated (Table 8.3).

Table 8.3. The Coding Triplets Obtained from Experiments with poly-AU (1 : 5)

Amino-acid residue	Calculated triplet fractions				Total fraction	Amino-acid incorporation
	3A	2A1U	1A2U	3U		
Asn	—	4	—	—	4	6.6
Ile	—	4	20	—	24	20
Leu	—	—	20	—	20	15
Lys	0.8	4	—	—	4.8	3.1
Tyr	—	—	20	—	20	25
Phe	—	—	—	100	100	100

In this way there were found the following triplets: 3U for Phe, 1A2U for Tyr, 2A1U and 1A2U for Ile, 1A2U for Leu, 2A1U for Asn, and 3A and 2A1U for Lys. It was subsequently found that the last-named triplet did not code for Lys; the rest were confirmed.

However, such experiments do not yet provide the complete deciphering of the code. It remains to be unknown which of the three codons AUU, UAU or UUA does code for Tyr, etc. In subsequent experiments Nirenberg replaced polynucleotides with trinucleotides of known structure. So, trinucleotide-tRNA-amino acid (aminoacyl) complexes are formed in the systems. The synthesis of a polypeptide does not occur in this case, but since the trinucleotide imitates the codon, the formation of the complex allows it to be read out. For this to be done, all the tRNA's that become consecutively bound to the labeled amino acids must be studied. All the 64 triplets were thus studied and the amino acids to which they became bound were identified. The final stage of deciphering, which confirmed these results, was conducted by Khorana (1966). He synthesized oligodeoxyribonucleotides which were recurring triplets of known structure [for example, $(TTC)_4$]. Such an oligomer was then used as a template for the *in vitro* synthesis of a DNA-like polymer in a system containing the nucleoside triphosphates of T, C, G, A, DNA polymerase and the ions required (cf. page 268). This involved the synthesis of a DNA-like double helix. The two chains contained complementary triplets which recurred n times. Then, each of the chains served as a template for transcription—for the synthesis of a polyribonucleotide with the aid of RNA polymerase. Thus, Khorana produced two chains which imitated mRNA with a known sequence of recurring triplets. Both chains were introduced into a cell-free system and the amino acid incorporation into the polypeptide fraction was determined by the Nirenberg method. These elegant experiments enabled six codons to be tested in a single multistep synthesis in accordance with the following scheme:



The polymer $(GAA)_n$ contains the codons GAA, AAG, and AGA, and the polymer $(UUC)_n$ contains the codons UUC, UCU, and CUU. Since the functionality of the homopolymers $(AAA)_n$, $(CCC)_n$, $(GGG)_n$, and $(UUU)_n$ has already been determined, in order to check

up the remaining 60 codons, one must carry out 10 such multistep experiments.

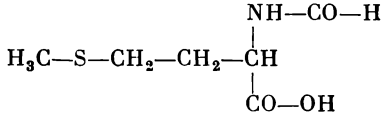
The thus established genetic code is shown in Fig. 8.12. Each RNA codon is designated *xyz*.

The codons GUG (Val) and AUG (Met) code for the amino acids in the middle of the protein chain. At the same time, they serve as

x \ y	A	C	G	U	z
A	Lys Asn Lys Asn	Thr	Arg Ser Arg Ser	Ile Ile Met Ile	A C G U
C	Gln His Gln His	Pro	Arg	Leu	A C G U
G	Glu Asp Glu Asp	Ala	Gly	Val	A C G U
U	— Tyr — Tyr	Ser	— Gys Trp Cys	Leu Phe Leu Phe	A C G U

Fig. 8.12. The table of the genetic code.

initiators of the chain synthesis, coding formylmethionine at its N-terminus end:



If the artificial synthetic template does not contain these codons, the chains are synthesized that start at an arbitrary unit and have therefore different N-terminal ends. On the contrary, in the presence of GUG and AUG there are formed standard F-Met terminal ends. Natural proteins, however, do not contain F-Met. In the cell-free system of *E. coli* there are formed chains with N-terminal ends F-Met-Ala, F-Met-Ser, but not F-Met-Met. Natural proteins of *E. coli* usually have Met, Ala, Ser at their N-terminal ends. One might think that in living systems two enzymes act—one of them

cleaves F-Met from the chain, and the other splits off the formyl group from Met.

The codons UAA, UAG, and UGA are terminal. They are responsible for the termination of the protein chain and do not code any amino acids.

The genetic code is universal, or very nearly so. Identical codons act in an analogous manner in different organisms. The universality of the code is directly proved by the multiplication of phages and viruses in cells. Virus DNA or RNA utilizes the biosynthetic machinery of the cell and synthesizes its own proteins. At the same time, it has been found that poly-U stimulates the incorporation of Phe into cell-free extracts from the cells of mammals and algae. The same refers to other synthetic polynucleotides. The trinucleotide technique of Nirenberg has been applied to cell-free extracts from the cells of the frog *Xenopus laevis* and of a guinea pig and led to identical results.

The sole known deviation of the code from universality has already been mentioned (see page 280). In the synthesis of cytochrome oxidase in the human mitochondria the codon AUA codes for Met rather than for Ile, and the codon UGA codes for Trp and not for chain termination. The code table is found in this case to be more symmetric in accordance with the prediction made by Ičas in 1969, ten years before this discovery.

The deciphering of the genetic code is the most significant achievement of molecular biology, biochemistry, and biophysics. The entire path from the setting-up of the problem to its solution took a little more than 10 years or so, a very short period of time.

The establishment of the code raised new problems. Does the genetic code have a physical, molecular meaning or is the correlation between the codons and amino acids accidental? What can be said in this connection about the origin and evolution of the code? How are point mutations associated with the specific features of the code table? What factors affect the reading of the code, the transcription and translation processes? The answers, far from complete, are given in the subsequent discussion. Here we shall confine ourselves to the physical meaning of the code.

First, attention should be focused on the nature of the degeneracy of the codons (Fig. 8.12). For 32 codons, i.e., for half of all codons, there is complete degeneracy with respect to the third nucleotide z. In other words, an amino acid is fully specified by the first two nucleotides x and y and is independent of the nucleotide z. For doubly degenerate codons the nature of the amino acid is different for $z =$ A, G (purine) and for $z =$ C, U (pyrimidine).

Sixteen doublets xy might be divided into two groups of eight in such a manner that the first octet would contain xy doublets

Table 8.4. The Codon System

First octet				Second octet				
x	y	n	Residue (z = A, G, U, C)	x	y	n	Residue (z = A, G)	Residue (z = U, C)
G	G	6	Gly	G	A	5	Glu	Asp
G	C	6	Ala	A	G	5	Arg	Ser
C	G	6	Arg	U	G	5	term.* (z = A) Trp (z = G)	Cys
C	C	6	Pro	C	A	5	Gln	His
C	U	5	Leu	A	A	4	Lys	Asn
G	U	5	Val	A	U	4	Ile (z = A) Met (z = G)	Ile
A	C	5	Thr	U	A	4	term.*	Tyr
U	C	5	Ser	U	U	4	Leu	Phe

* Term. = "nonsense" triplet (chain termination).

coding for the amino-acid residue, irrespective of z; and the second octet will contain the remaining doublets (Rumer). This is shown in Table 8.4.

The columns n give the number of hydrogen bonds between the nucleotides of the xy codon and the complementary nucleotides of the $x'y'$ anticodon. The doublets xy of the first and second octets differ sharply in composition. The first octet contains A only once, and in the second there is only one C. In the first octet, for both x and y (G + C)/(A + U) = 3, and in the second (G + C)/(A + U) = = 1/3. The value of n, which may be called the degree of complementarity, is 6 and 5 ($\bar{n} = 5.5$) in the first octet and 5 and 4 ($\bar{n} = 4.5$) in the second. It may be expected that the greater is the value of n, the weaker is the interaction z-z' between codon and anticodon, since the interaction xy-$x'y'$ is sufficiently stable.

The degeneracy with respect to z in the first octet is associated with the so-called wobbles of the z-z' interaction (Crick). This interaction, unlike xy-$x'y'$, is not unambiguous, as evidenced by Table 8.5.

Table 8.5. The Formation of Pairs by the Third Nucleotide of the Anticodon

z'	A	U	G	C	I
z	A	A, G	C, U	G	A, C, U

The letter I stands for the minor nucleotide inosine.

The primary structure of proteins is coded genetically and the three-dimensional structure of the globule is coded in a biologically functional manner. The mutation replacements of the template nucleotides affect the functionality of the protein in various ways. We have seen that the hydrophobicity of the amino-acid residue is very important to the structure of the globule (see Sec. 4.5). Accordingly, mutations that strongly alter the hydrophobicity of the residue must have a stronger effect on the biological properties of the protein than mutations that alter the hydrophobicity only slightly. Mutations of the first type are more detrimental to the existence of the species and individual than the second. One might think that the evolution that has led to the existing code progressed in the direction of increasing *reliability* of the code in the sense that the fraction of more deleterious mutations diminished. Let us see how reliable the code is.

Let us inspect Table 4.5 (page 118) which gives the hydrophobicities of the amino acids. The average difference in these values upon arbitrary replacement of one amino acid by another is 5370 J/mole. Within the conditional classification of amino acids into two classes—the hydrophobic (the first ten) and the hydrophilic class (the second 10 amino acids), the average differences are 3370 and 1640 J/mole, respectively.

The class of amino acids depends primarily on the second nucleotide y in the xyz codon. With $y=$U the residue is always hydrophobic (Ile, Met, Leu, Val, Phe). Let us call mutations that do not alter the class of the amino-acid residue correct and those which do alter the class incorrect. The latter are more deleterious to the structure of the protein. Replacement of one of the nucleotides in the codon xyz gives rise to the following distribution of correct and incorrect mutations:

	120 correct	74 correct	156 correct
Replacement of	x	y	z
	54 incorrect	102 incorrect	20 incorrect

there are altogether 350 (66.5 percent) correct replacements and 176 (33.5 percent) incorrect replacements. The correct replacements include "*silent*" *mutations*, i.e., those cases where the replacement of a nucleotide is not accompanied by the replacement of the residue because of the degeneracy of the code. The code is built so that the probability of a correct mutation is twice as high as the probability of an incorrect mutation.

Let us calculate the average differences in the hydrophobicities of the residues that correspond to single replacements of nucleotides. We obtain:

upon replacement of x 4200, y 5370, z 1420 J/mole

and, on the average, 3660 J/mole for all the three nucleotides, which is much less than 5370 J/mole upon accidental replacement of the amino acid. Experiment gives the average difference in the hydrophobicities of the original and replacing amino acids for 70 mutants of human hemoglobin 3500 J/mole and for 6 cytochromes *c* 3780 J/mole, etc. The analysis of 423 replacements by comparing six homologous proteins of different species (cytochrome *c*, hemoglobins α and β, insulins *A* and *B*) gives an average hydrophobicity difference of 3240 J/mole.

Thus, the code is highly (though not absolutely) reliable with respect to correct mutations, providing their higher probability. The advantage of correct mutations is due to the three-dimensional structure of the protein in an aqueous environment and, hence, to the specific physical properties of water. The correlation between codons and amino acids is found to be dictated by the physics of water.

8.7. Mutations

Mutations occur either spontaneously (spontaneous mutations) or under the influence of external factors—chemical or radiation effects on chromosomes and genes (induced mutations). The following two types of mutations should be distinguished: *chromosomal mutations*—reconstitutions of chromosomes observable under the microscope, and *point mutations* (or replacement or gene mutations). The former are changes in the supermolecular structures, and the latter are changes in the nucleotide sequence in DNA and, accordingly, in mRNA. Here we shall consider point mutations.

There exist four types of point mutation.

(a) Mutations leading to replacements of nucleotides in codons and altering the sense of the codon, i.e., the amino acid residue in the protein. Such mutations are called *missense mutations.*

(b) The replacements in codon leading to the codon degenerated together with the original one, i.e., which do not change the amino-acid residue in the protein. These *silent mutations* do not manifest themselves phenotypically, but they are very important to evolution (see Sec. 17.9).

(c) Mutations in which a normal codon that specifies an amino acid is changed to one of the three terminal codons: UAA, UAG, UGA. Such mutations are known as *nonsense mutations.* They lead to the termination of the protein chain and are especially detrimental.

(d) *Frameshift mutations* which result in a deletion or an insertion of a nucleotide. They are so called because they shift the reading frame of the mRNA triplet codons (see page 277).

Spontaneous gene mutations are determined by errors in DNA replication that arise because of the thermal motion of atoms and

molecules. Evidently, these errors of transcription and translation are not inherited.

The insertion of a nucleotide that is not complementary to the template nucleotide, the deletion or replacement of a nucleotide usually lead to the formation of a loop in the DNA double helix (see page 245). In subsequent replications the DNA loop disappears as a result of the semiconservative synthesis, but the primary structure of DNA remains unchanged. The formation of loops may be accompanied by the formation of a pair different from a Watson-Crick pair due to the ability of a nitrogen base to form unusual hydrogen bonds (page 248) and also due to tautomerism (page 39).

The replacement of the pair A-T by G-C is evidently thermodynamically favourable, since G is bound to C more strongly than is A to T. If it were for thermodynamics alone, the relative G-C content in DNA should have increased during the evolution. This is not the case in actuality—in higher organisms the G-C content is stabilized at the 40-45 percent level. The evolutionary formation of a "G + C organism" is biologically meaningless because triplets that do not contain A and T (A and U in mRNA) code only for Pro, Arg, Ala, and Gly, i.e., only one-fifth of all the amino-acid residues.

Different point mutations have different probabilities. Vogel and Kopun (1977) analyzed 1293 amino-acid replacements in phylogenetically linked proteins which are accounted for by single replacements of nucleotides xy in the codons xyz. The replacements of z are omitted because out of $61 \times 3 = 183$ replacements in sense codons 7 replacements lead to chain-termination codons and 126 replacements are silent mutations. The numbers n_i of replacements $x \to x'$ and $y \to y'$ in RNA codons found in this way are listed in Table 8.6.

Table 8.6. The Numbers of Mutation Replacements (n_i) $x \to x'$ and $y \to y'$ in RNA Codons for Phylogenetically Linked Proteins

x, y \ x', y'	A	G	C	U
A	—	237	114	60
G	272	—	92	71
C	96	73	—	92
U	54	63	69	—

In order to find the relative probabilities of such mutations, one should take into account certain differences between the numbers of nucleotide replacements m_i that may be observed in the form of amino-acid replacements according to the code. We thus obtain Table 8.7.

Table 8.7. The Numbers of Essentially Observable Replacements (m_i) $x \to x'$ and $y \to y'$ in Codons

x, y \ x', y'	A	G	C	U
A	—	30	28	27
G	30	—	31	28
C	28	31	—	27
U	27	28	27	—

The relative probabilities of replacements are found from the formula

$$p_i = \frac{n_i}{\sum\limits_i n_i} \frac{\sum\limits_i m_i}{m_i} \tag{8.1}$$

Normalizing p_i to unity, we obtain Table 8.8.

Table 8.8. The Normalized Relative Probabilities of the Replacements $x \to x'$ and $y \to y'$ in RNA Codons on the Basis of Data for Proteins

x, y \ x', y'	A	G	C	U
A	—	0.177	0.091	0.050
G	0.201	—	0.067	0.057
C	0.076	0.053	—	0.076
U	0.045	0.051	0.057	—

The replacements of nucleotides can be divided into two groups: transitions and transversions. *Transitions* are point mutations in either DNA or RNA in which there is a replacement of one purine by another or a replacement of one pyrimidine by another: A $\rightleftharpoons$ G or C $\rightleftharpoons$ U; *transversions* are point mutations in either DNA or RNA in which there is a replacement of a purine by a pyrimidine or vice versa: A $\rightleftharpoons$ C, A $\rightleftharpoons$ U, G $\rightleftharpoons$ C, G $\rightleftharpoons$ U. According to Table 8.8, the probability of transitions $p = 0.68$ and that of transversions $1 - p = 0.32$. The data referring to 198 mutants of human hemoglobin are somewhat different: $p = 0.62$ and $1 - p = 0.38$.

These data characterize two circumstances: the true probabilities of nucleotide replacements in DNA and RNA and the survival of the corresponding mutant proteins. Data have recently been obtained that refer directly to RNA, namely, 247 replacements. The corresponding normalized relative probabilities are presented in Table 8.9.

Table 8.9. The Normalized Relative Probabilities of Replacements in RNA

x, y \ x', y'	A	G	C	U
A	—	0.065	0.040	0.053
G	0.117	—	0.109	0.121
C	0.093	0.113	—	0.121
U	0.049	0.028	0.089	—

The excess of transition probabilities over transversion probabilities is retained, but it is not too large: $p = 0.56$ and $1 - p = 0.44$.

There is a considerable difference in the replacement probabilities found from the primary structures of proteins (Table 8.8) and directly from the primary structures of RNA (Table 8.9). In the first case, the probabilities of transitions A $\rightleftharpoons$ G by far exceed the probabilities of transversions and transitions C $\rightleftharpoons$ U, whereas in the second this is not observed. Hence, the replacements A $\rightleftharpoons$ G give a smaller fraction of lethals than the other replacements. This is accounted for by the fact that the amino-acid residues connected by transitions A $\rightleftharpoons$ G are substantially closer to each other than those connected by other transitions. Indeed, the transitions A $\rightleftharpoons$ G produce the smallest changes in hydrophobicity of amino-acid residues. On the contrary, the hydrophobicity changes that occur in transitions C $\rightleftharpoons$ U are twice as large. Table 8.10 gives the averaged changes in the hydrophobicities of amino-acid residues upon nucleotide replacements in mRNA.

Table 8.10. The Average Differences in the Hydrophobicity of Amino-Acid Residues (in kJ/mole) for Replacements $x \rightarrow x'$ and $y \rightarrow y'$ in RNA Codons

x, y \ x', y'	A	G	C	U
A	—	2.94	4.97	5.10
G	2.94	—	4.35	6.24
C	4.97	4.35	—	5.69
U	5.10	6.24	5.68	—

Thus, the mutations A $\rightleftharpoons$ G are the least detrimental to the structure and functions of the protein and have the least lethal consequences. The true probabilities of replacements are presented in Table 8.9. They are in agreement with quantum-mechanical calculations of the interaction energies in various nitrogen base pairs.

We have considered missense mutations which lead to the replacement of an amino-acid residue. We shall now turn to nonsense mutations which lead to the termination of the protein chain. There are altogether 23 such mutations, which involve single nucleotide replacements in codons *xyz*: 9 replacements in *x*, 7 replacements in *y*, and 7 replacements in *z*. Of these 23 mutations 18 are transversions and only 5 are transitions. The lower probability of transversions lowers, to a certain degree, the probability of a lethal nonsense mutation.

The values of the probabilities of point mutations in codons given in Table 8.9 allow one to calculate the relative probabilities of chain-termination (nonsense) mutations. These values are listed in Table 8.11.

Table 8.11. The Relative Probabilities of Mutations Leading to Terminal Codons UAA, UAG, and UGA

1.	U → A in codons UU**G**, UG**U**	0.049
2.	A → U in codons A**A**A, A**A**G, A**G**A	0.053
3.	U → G + U → A in codons U**U**A, UA**U**	0.077
4.	C → A in codons U**C**G, UG**C**	0.093
5.	G → U in codons **G**AA, **G**AG, **G**GA	0.121
6.	C → U in codons **C**AA, **C**AG, **C**GA	0.121
7.	C → G + C → A in codons U**C**A, UA**C**	0.206
8.	2 mutations G → A in codons UG**G**	0.234

Thus, the nonsense mutations in the codon UGG, which codes for Trp, are the most probable. Of the 9 mutations of this codon 2 are nonsense mutations which occur with an increased probability. It may be supposed that this circumstance is responsible for the fact that in a gene that codes Trp-tRNA there occur mutations that convert this tRNA to a suppressor RNA. The intactness of the codon UGG is especially important.

As regards silent mutations, their role in evolution is determined not by amino-acid replacements but by the fact that the jointly degenerate codons may mutate differently. We shall give one example. The mutation UUG (Leu) → UUA (Leu) does not change the amino-acid residue. But the probability of the nonsense mutation of UUA is 0.077, and that of the nonsense mutation of UUG is equal to 0.049 (Table 8.10). See also Sec. 17.9.

8.8. Regulation of Genes

The functioning of the genes, i.e., protein biosynthesis, undergoes delicate regulation in living systems. The regulation at the molecular level is brought about by phenomena of molecular recognition accomplished via weak interactions. In the final analysis, it is exactly molecular interactions, which produce feedback required for regulatory processes, that determine the entire path of the biological development of the cell. We have already discussed feedback phenomena at the molecular level—the phenomena of allosterism (page 221).

The synthesis of the protein is distorted by the action of mutagenic factors on DNA. Nonmutagenic substances can also exert a substantial effect on the operation of genes. The *induced synthesis of enzymes* has been discovered in prokaryotes. The wild-type *E. coli* cells which grow in a medium with a nongalactoside source of carbon do not practically synthesize the enzyme β-*galactosidase*, which catalyzes the hydrolysis of the galactoside lactose. Addition of a substance capable of acting as an inducer of synthesis, say methylthiogalactoside, to the growing culture of the wild-type *E. coli* increases the rate of synthesis of the galactosidase 10^3 times. The wild-type *E. coli* is an *induced type*. At the same time, there are also mutant strains of *E. coli* that synthesize β-galactosidase in the absence of an inducer. These mutants are called *constitutive* mutants. An inducer affects the genetic system of the cell. Jacob and Monod (1961) carried out a genetic analysis of an induced synthesis on the basis of a simple hypothesis which was later convincingly confirmed. The induced synthesis is suppressed by a specific substance, called a *repressor*, present in the cytoplasm. The repressor is synthesized by a special *regulator gene*. It was later demonstrated that repressors are proteins. The repressor acts at the *operator gene* which controls the transfer of information from several structural genes to the proteins being synthesized. The action of the repressor stops the operation of the entire set of these genes—the entire *operon*.

The inducer of synthesis interacts with the protein-repressor and blocks ("turns off") its effect on the operator gene.

Figure 8.13 shows the scheme of the genetic system described above. In the case under consideration the repressor controls the synthesis of at least two enzymes: β-galactosidase and β-galactoside permease. The second enzyme determines the rate of incorporation of β-galactosides into the bacterial cells through the membranes. The syntheses of the two enzymes are properly correlated.

The operon model has been verified by genetic investigations, which have proved the existence of a regulator gene and an operator gene and the occurrence of their mutations. The synthesis of β-galactosidase in *E. coli* is controlled by the so-called *lac operon*. This operon

is acted on by a *lac repressor*—a tetrameric protein with a molecular mass of 150,000.

Systems are known to exist in which one specific repressor acts on several different operons.

In the regulation of the operon as well as in the processes of DNA replication and transcription, which are catalyzed by the respective polymerases, we encounter the processes of protein-nucleic acid recognition which are very important to biology and biophysics.

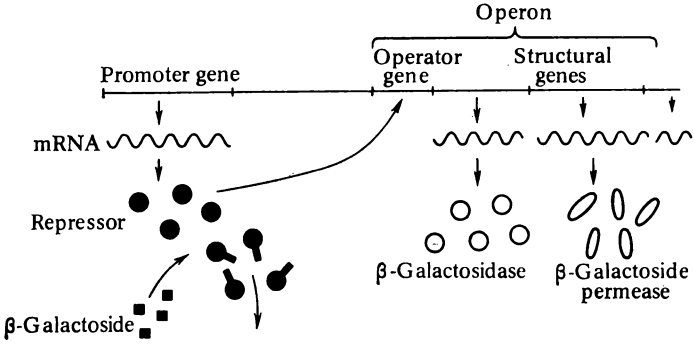


Fig. 8.13. The operon model.

Here the "hypercycle" is realized (Eigen; see Sec. 17.4): DNA and mRNA are responsible for protein synthesis, which in turn governs the functioning of the nucleic acid.

Gursky, Gottikh and their colleagues have proposed a code of nucleic acid-protein recognition which determines the regulation of transcription (1976). It is assumed that the region of the regulatory protein consists of two antiparallel segments of the polypeptide chain which form the β-structure. The recognition is based on the complementarity of this structure and the base-pair sequence of DNA. An important property of this sequence is the asymmetric distribution of guanines between the two strands of DNA. In the code proposed, the six amino-acid residues—Ser, Thr, Asp, His, Gln, Cys—and their sequence in the stereospecific region of the protein determine the base-pair sequence in DNA, to which the given protein becomes predominantly bound. The code worked out on the basis of stereochemistry is confirmed by the interaction between the lac repressor and the lac operon and by other examples.

The control of the operon by the operator gene is presumably determined by the fact that the synthesis of mRNA starts at the end of the operon that adjoins the operator gene.

The above-described method of control and regulation of protein biosynthesis in prokaryotes does not yet provide for the regulatory needs of the cell. The proteins that are coded for by the same operon may be needed in different amounts and at different times. In order to get an insight into the corresponding regulatory phenomena, a detailed examination of the transcription process is required. It is reasonable to consider the onset of the synthesis of RNA on DNA (initiation) and the continuation of this synthesis (elongation) as two independent steps.

The DNA-dependent synthesis of RNA is suppressed by a number of antibiotics, such as actinomycin, which block the DNA template. Antibiotics of the rifamycin group act, on the contrary, on the RNA polymerase. Rifamycin acts at the initiation stage, inhibiting the formation of the first internucleotide bond. Other substances that affect the polymerase inhibit the elongation.

The regulation of transcription is realized not in all cases by the Jacob-Monod scheme via a negative regulatory repressor factor. In the case of phage DNA's the readout of definite genes does not occur in the absence of any repressors either. For these genes to be switched on, positive regulatory factors are required.

On infecting the cell of *E. coli* with T-even phages there is realized an accurately regulated time sequence of transcription processes. In several minutes after infection the synthesis of mRNA and of the proteins of the host cell is turned off and several new enzymes needed for the synthesis of phage DNA are synthesized, this being followed by the synthesis of the structural proteins of the phage. The specific phage mRNA's appear not immediately but consecutively (Khesin and coworkers, 1962, 1963). At the early and final stages of the development of phages T2 and T4 there are formed, in the cell of *E. coli*, various sets of mRNA's synthesized on various groups of genes. The appearance of various groups of mRNA depends on the processes of protein synthesis and replication of phage DNA. The models proposed to account for these phenomena are based on the consideration of the complex subunit structure of RNA polymerase, which undergoes changes under the influence of protein regulatory factors. Detailed molecular mechanisms of the time regulation of protein synthesis are so far unknown.

This regulation is evidently effected at all the stages of protein synthesis. The operation of polymerases, aminoacyl-tRNA-synthetase and ribosomes is regulated. It has been found that antibiotics affect the translation of the code by exerting an influence on the ribosomes. Streptomycin, which upsets translation in a cell-free system as well, is incorporated into the 30S subunits of the ribosomes.

Cyclic adenosine monophosphate (cAMP, see page 44) stimulates the operation of genes at the level of transcription and serves as a chemical trigger of transcription. Transcription begins when the com-

plex of cAMP with the receptor-protein activates some promoter site of DNA at the beginning of the operon. RNA polymerase binds to the activated promoter site and then moves along the DNA chain, organizing the synthesis of mRNA. The cAMP-receptor complex does not contribute to transcription in the presence of a specific repressor protein. Cyclic AMP can stimulate the transcription of a number of various operons.

The gene regulation in *eukaryotic* cells is naturally much more complicated. The genes of eukaryotes are present in *chromosomes*—supermolecular structures which are *nucleoproteins.* Nucleoproteins are organized DNA-protein complexes.

All somatic cells of a particular multicellular organism contain the same set of genes identical with the genome of the original zygote. (We distract from somatic mutations.) At the same time, the cells of various tissues differ from one another both morphologically and functionally. Their differences are reduced to the following: different proteins function in different cells of the same organism. This means that different genes operate in different cells and the molecular significance of *cell differentiation* consists in the regulation of gene operation. In a given species of cells the translation is only effected for a small fraction of genes present.

Proteins that enter into the composition of chromatin and chromosomes are subdivided into *histones* and *nonhistone proteins* (NHP). The following five varieties of histones, which are small basic proteins containing large amounts of Arg and Lys, are known to exist: H1, H2A, H2B, H3, and H4. The primary structures of histones isolated from various organisms have been established. This structure is exceedingly stable during evolution for nearly all the five types of histones. For example, histones H2 from peas and calf thymus differ in only two amino-acid residues of the 102 present in the molecule. The constancy of the primary structure of histones is presumably associated with the fact that the entire histone molecule is functional: some of its portions are responsible for binding to DNA through the salt bridges Arg, Lys-phosphate, and the others for interprotein interactions. *Chromatin* is built up of tightly packed *nucleosomes*—fragments 10 nm in diameter, into which chromatin is split by the action of *nucleases.* There are three types of mononucleosomes: MH1 which contains 145 base pairs of DNA and all histone fractions (two molecules of each), except H1; MH2 which contains 160-170 base pairs, the same 8 histone molecules as MH1 and one H1 molecule; and MH3 which contains 200 base pairs and the same 9 histone molecules as MH2. According to Crick and Klug, DNA in nucleosomes is coiled around the protein into a superhelix; this superspiralization is determined not by a uniform folding of the double helix but by its abrupt bending through about every 20 base pairs (1975). The coiling of DNA into a superhelix is evidently brought about by

neutralizing the negative charges of the phosphate groups by the positive charges of the histones. The presence of 8 or 9 different histones in nucleosomes makes possible for nucleosomes to have a large variety of structures. Tsanev and Sendov presumed the existence of a specific code for gene blocking, assuming that the coding is determined by the different combinations of five histones (1966).

During the course of cell development the conformations of histones and nonhistone proteins and their DNA complexes undergo change and the genome experiences functional changes and becomes more or less accessible to the action of the regulatory proteins of cytoplasm. At the giant chromosomes of two-wing insects (*Diptera*) there appear, at a certain stage of development, chromosomal *puffs*—swollen portions, which are the loci of the most intensive synthesis of RNA. The histones at these sites undergo chemical and conformational changes, which is what provides a change in the functionality of the corresponding genes. Presumably, the histones in the puffs are less strongly bound to DNA; they are more accessible to the action of proteases and are more easily isolated. Accordingly, the histones in the puffs do not interfere with the operation of RNA polymerase. Under normal conditions the histones impede transcription.

Georgiev has studied the mechanism of inhibition of RNA synthesis by histones, using the method of the double-label technique. The ^{32}P-phosphate-labeled ATP and GTP were used to determine the initiation of the RNA chain, and ^{14}C-UTP was employed for the determination of the overall rate of growth of the RNA chain. The addition of histones diminished the $^{14}C/^{32}P$ ratio in the synthesized RNA. This shows that either the rate of chain growth is slowed down or relatively short chains are formed, i.e., the histones impede the movement of the polymerase along the template. The template activity of chromatin is one order of magnitude lower than that of chromatin in free DNA. The removal of histones from chromatin increases its template activity.

Nonhistone proteins contain acidic amino-acid residues and not basic ones. They are highly heterogeneous. Their molecular masses range from 10,000 to 150,000. They are functionally different. The properties and structures of nonhistone proteins have been studied little, but their participation in the regulation of genes is undoubtful. The ability of nonhistone chromatin proteins to stimulate the synthesis of RNA in cell-free systems depends on the state of their phosphorylation. A hypothesis has been worked out, according to which the gene is incorporated through addition of a nonhistone protein to a histone-repressed specific site of DNA. The nonhistone proteins are phosphorylated and become negatively charged. Therefore they repel the negatively charged DNA and leave it together with the positively charged histones. There is left a free portion of DNA capable of transcription.

The genomes of eukaryotes are very complex. They contain repetitious genes and inverted genes (*palindromes*). The genetically functional portions of DNA alternate with nonfunctional portions (*spacers*). There are transcribable and nontranscribable spacers. It has been recently found that even within the genes, i.e., the functional portions of DNA, there are intervening nonfunctional portions called *introns* in this case. Georgiev and his coworkers have found that eukaryotes have gene assemblies that form whole regulatory systems—*transcriptons* (Fig. 8.14). Their transcription gives rise to

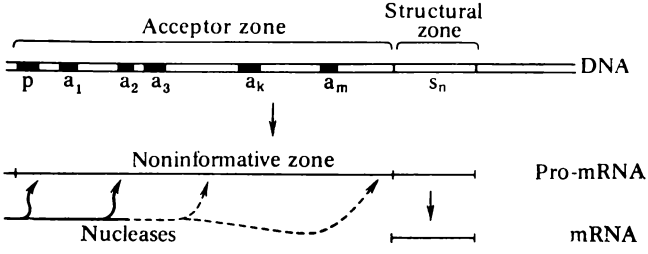


Fig. 8.14. Schematic representation of the transcripton: p—promotor; $a_1 \ldots a_m$ are acceptor sites; s_n are structural genes.

pro-mRNA, only a small part of which is a true mRNA responsible for protein synthesis. The greater part of *pro*-mRNA is noninformational and degrades under the action of nucleases. The breakdown of this part of *pro*-mRNA and the separation of the true mRNA, shown schematically in Fig. 8.14, are called *processing*.

Thus, the transcripton consists of a very large genetically nonfunctional zone of DNA, which may be called the acceptor zone, and a small structural zone. It is exactly the acceptor zone that contains portions that interact with regulatory proteins. At the beginning of the transcripton, whose size is much larger than that of the bacterial operon, there is a promoter region to which RNA polymerase attaches itself. The movement of the polymerase along the transcripton length (transcription) is regulated by the interaction of regulatory proteins with acceptor regions.

The transcripton model provides a much subtler regulation of gene expression, blocking and activation than the regulation of genes in prokaryotes. The interaction of NHC proteins and histones with acceptor regions affects the structure of chromatin and chromosomes, determining the packing of DNA, on which the possibilities of transcription depend. A single transcripton contains various acceptor regions which recognize different regulatory proteins. And,

conversely, different transcriptons can recognize the same protein. Accordingly, the addition or separation of one protein can turn on or off the entire assembly of transcriptons.

The large size of the DNA acceptor zone turns out to be necessary precisely for a fine and diverse regulation, without which no differentiated multicellular organism would be able to exist. According to the transcripton model, mRNA is at the end of the giant *pro*-mRNA. Since there are acceptor regions which are identical for different transcriptons, the initial parts of *pro*-mRNA must contain identical sequences for different *pro*-mRNA. This has been confirmed by experiment.

Orgel and Crick (1980) formulated the hypothesis of "*selfish DNA*". The greater part of DNA in cells of higher organisms is not functional—it has no effect on the phenotype. It might be expected that the "selfish" DNA is replicated independently, behaving in this sense like a not too harmful parasite. The presence of an excess amount of such DNA can, however, slow down the rate of meiosis. The size of the genome sets up a limit, beyond which the development cannot be speeded up.

The problems of the regulation of gene expression in eukaryotes have only recently become the concern of biophysics. The complex nucleoprotein systems of chromatin and chromosomes have not been studied sufficiently well for a clear-cut formulation of many physical problems to be made possible. At the same time, the physico-mathematical modelling of regulatory processes of development is successfully progressing (see Chapter 17).

Chapter 9 Nonequilibrium Thermodynamics in Biology

9.1. Information and Entropy

While studying the structure and properties of biopolymers, we have so far, as a rule, confined ourselves to the consideration of equilibrium states. As far as a developed structure is concerned, this is quite natural. But the emergence of a structure, i.e., biosynthesis and self-assembly, is a series of nonequilibrium processes which proceed irreversibly. The more so it refers to the functioning of a biological system as a whole, to the processes of evolutionary and individual development.

As we have seen, a living organism is a kind of chemical machine that works due to molecular feedforward and feedback relationships. Molecular signalling serves to convey informational messages. Accordingly, a general physical treatment of biological systems is based on information theory which plays an important part in thermodynamics.

Information theory introduces a measure of the *amount of information.* Suppose that there are P_0 different, equally probable events. For example, in coin tossing $P_0 = 2$ and in die casting $P_0 = 6$. The larger the value of P_0, the greater is the uncertainty until we obtain information about the event and the larger is the amount of information (we shall call it simply information in the subsequent discussion) when we obtain the message. In the initial situation, before the coin or die is thrown, the information is equal to zero, $I_0 = 0$, and in the final situation, $I_1 \neq 0$. The measure of information must evidently be connected with P_0. It will be natural to demand that information be additive for independent events (say, in throwing two dice).

Thus, if there are two sets of events, P_{01} and P_{02}, so that the total number of events is

$$P_0 = P_{01}P_{02} \tag{9.1}$$

(in tossing two dice $36 = 6 \times 6$), then we have

$$I\,(P_{01}P_{02}) = I\,(P_{01}) + I\,(P_{02}) \tag{9.2}$$

This relation is satisfied by the only solution possible:

$$I = K \ln P_0 \tag{9.3}$$

the constant K being arbitrary. The base of the logarithm is therefore also arbitrary. It is convenient to use the binary system and logarithms to the base 2. If we form all possible "words" or sequences of two digits 0 and 1 and of length n, there will be $P = 2^n$ possibilities. Let us require that

$$I \equiv K \ln P = Kn \ln 2 = n \tag{9.4}$$

that is,

$$K = 1/\ln 2 = \log_2 e \tag{9.5}$$

and

$$I = \log_2 P \tag{9.6}$$

Here the information is measured in *bits*. Thus, at $P_0 = 2$ and $P_1 = 1$ (the tossing of a coin)

$$I = \log_2 P_0 - \log_2 P_1 = 1 - 0 = 1 \text{ bit}$$

How many bits does a three-digit number contain? The first digit has 9 different values—from 1 to 9, the second and third—10 values each, from 0 to 9. We have

$$I = \log_2 9 + 2 \log_2 10 = 9.28 \text{ bits}$$

The definition of information adopted corresponds to the binary system in which any number is written as powers of the number 2 by means of the digits 0 and 1. One decimal unit gives 3.32 bits, i.e., the binary notation of the number requires, on an average, 3.32 times more figures than does the decimal notation.

Suppose we have a message containing N successive cells—a text of N letters. Each of N cells may contain one of M letters (in the Russian language $M = 32$). The message consists of N_1 letters A, N_2 letters Б, etc. We have

$$N = \sum_{j=1}^{M} N_j \tag{9.7}$$

The probability of occurrence of a particular letter is

$$p_j = N_j/N, \quad j = 1, 2, \ldots, M \tag{9.8}$$

with

$$\sum_{j=1}^{M} p_j = 1 \tag{9.9}$$

The total number of various sequences of N letters of the M-letter language, i.e., the number of various possible messages, is equal to

$$P = \frac{N!}{\sum_{j=1}^{M} N_j!} \,. \tag{9.10}$$

The information in a single message is equal to (we use Stirling's formula which gives good results at $N_j > 100$)

$I = K \ln P =$

$$= K\left[\ln(N!) - \sum_{j=1}^{M} \ln(N_j!)\right] \approx K\left(N \ln N - \sum_{j=1}^{M} N_j \ln N_j\right) \tag{9.11}$$

or

$$I = -KN \sum_{j=1}^{M} p_j \ln p_j \tag{9.12}$$

and the information per letter is

$$i = I/N = -K \sum_{j=1}^{M} p_j \ln p_j \tag{9.13}$$

We have just obtained the *Shannon formula*—a more general expression for information which corresponds to a sequence of events that have nonequal probabilities p_j. Here, if $K = 1/\ln 2$, the information is expressed in bits, and if $K = \mathrm{k} = 1.38 \times 10^{-23}$ J/K (Boltzmann's constant), then I is expressed in J/K, i.e., in entropy units. We call the quantity

$$S = -\mathrm{k} \sum_{j} p_j \ln p_j \tag{9.14}$$

the *entropy*. We shall see later that this quantity is actually the physical entropy.

Let us show that a change in uncertainty leads to a gain in information. We pass over from the probability distribution $P = (p_1, p_2, \ldots, p_M)$ to the distribution $Q = (q_1, q_2, \ldots, q_M)$. But how does the information change? When the probability of an event changes from p to 1, the change in the information is equal to

$$\Delta I = \mathrm{k} \ln \frac{1}{p} = -\mathrm{k} \ln p$$

and with the change $p \to q$

$$\Delta I = \mathrm{k} \ln \frac{q}{p} \tag{9.15}$$

When the entire probability distribution is changed from P to Q, the change of the information is equal to the sum of the partial

changes ΔI multiplied by the finite probabilities q_j:

$$\Delta I = \mathrm{k} \sum_j q_j \ln \frac{q_j}{p_j} \tag{9.16}$$

This quantity is always positive; only at $Q = P$ is the change in information ΔI equal to zero. Let us prove this. At any x, except at $x = 1$,

$$\ln x > 1 - 1/x$$

Hence,

$$\sum_j q_j \ln \frac{q_j}{p_j} > \sum_j q_j \left(1 - \frac{p_j}{q_j}\right) = \sum_j q_j - \sum_j p_j = 0$$

Thus, the knowledge of the change $P \to Q$ diminishes the uncertainty and gives a positive gain in information.

Let p_j denote the probabilities of the system being in states with energy E_j. We have

$$E = \sum_j p_j E_j \tag{9.17}$$

We seek the maximum of the quantity S/k [Eq. (9.14)] with the simultaneous fulfilment of the conditions (9.9) and (9.17). We make use of the method of Lagrange multipliers. We multiply Eq. (9.9) by $\alpha - 1$ and Eq. (9.17) by β, add these expressions to Eq. (9.14) and vary the sum obtained. Equating the variation to zero, we get:

$$\delta \left[\frac{1}{\mathrm{k}} S - (\alpha - 1) \sum_j p_j - \beta \sum_j p_j E_j \right] = 0 \tag{9.18}$$

Differentiating with respect to p_j and setting the derivative to zero, we find that

$$-\ln p_j - 1 - (\alpha - 1) - \beta E_j = 0$$

from which

$$p_j = \exp(-\alpha - \beta E_j) \tag{9.19}$$

Substituting Eq. (9.19) into Eq. (9.14), we find the maximum entropy:

$$\frac{1}{\mathrm{k}} S_{\max} = \alpha \sum p_j + \beta \sum p_j E_j = \alpha + \beta E \tag{9.20}$$

On the other hand, from Eqs. (9.9) and (9.19) it follows that

$$1 = \sum_j p_j = e^{-\alpha} \sum_j e^{-\beta E_j} = e^{-\alpha} Z$$

that is,

$$e^{\alpha} = Z, \quad \alpha = \ln Z \tag{9.21}$$

Equation (9.20) yields

$$E - \frac{1}{k\beta} S_{max} = -\frac{1}{\beta} \ln Z \tag{9.22}$$

This is the well-known equation of thermodynamics and statistical physics. Here

$$Z = \sum_j \exp(-\beta E_j) \tag{9.23}$$

is the *partition function*; S_{max} is the equilibrium entropy; $\beta^{-1} = kT$, where T is the absolute temperature. Expression (9.19) is the *Boltzmann distribution function*:

$$p_j = \frac{\exp(-E_j/kT)}{\sum_j \exp(-E_j/kT)} \tag{9.24}$$

The average energy of the system, E, and the equilibrium entropy S (the subscript *max* is omitted) are functions of the volume V. Expression (9.22) gives the *Helmholtz free energy*:

$$E - TS = -kT \ln Z = F \tag{9.25}$$

The *Gibbs free energy* is defined as

$$G = F + pV = E + pV - TS = H - TS \tag{9.26}$$

where H is the enthalpy and p is the pressure.

Thus, the informational entropy (9.14) really coincides with the thermodynamic entropy. This has a real physical meaning—the information obtained must be paid for by an increase in entropy. Any measurement is associated with an increase in the entropy of the environment. If the information and entropy are measured at the same level of reception, the law of conservation holds:

$$I + S = \text{const} \tag{9.27}$$

The two quantities are measured in the same units. This means that the entropy is a measure of the deficiency of information about the system. The entropy S decreases with increasing I and vice versa. The quantities I and S have here a macroscopic meaning. For example, when a liquid is evaporated, the macroscopic information about the location of molecules in a certain part of space disappears. Accordingly, the entropy increases. The macroscopic information is converted to microscopic information, say to the information about the correlations between the velocities of molecules in the gas which appear upon their collisions. But this is no longer information—it can neither be recorded nor memorized. This is the entropy.

The equivalency of information (in bits) and entropy (in J/K) is similar to the equivalency of mass and energy in accordance with

the Einstein law:

$$E = mc^2$$

In our case, 1 bit of information is equivalent to $k \ln 2 = 10^{-23}$ J/K, i.e., to a very small thermodynamic value. On the entropy scale, the information "costs" very little. Nonetheless, it must be paid for. Let us evaluate the "price" for one bit of information in work units.

Suppose that we have an ideal gas containing N molecules at pressure p and temperature T. As a result of fluctuations, the volume of the gas decreases from V to $V - \delta V$. The work expended for such a decrease is $W = p\delta V$. We calculate the gain of information. Each molecule resides in volume V with a probability of 1 and in volume $V - \delta V$ with a probability of $1 - \delta V/V$. For N molecules the probability is equal to $(1 - \delta V/V)^N$. The gain in information is

$$\Delta I = K \ln \frac{1}{(1-\delta V/V)^N} \approx KN \frac{\delta V}{V} \quad (\text{at } N \gg 1)$$

Hence,

$$\frac{W}{\Delta I} = \frac{p\delta V}{KN\delta V} V = \frac{pV}{KN} = \frac{kT}{T}$$

The work expended per one bit of information is proportional to the temperature at which the information is determined. If ΔI is evaluated in bits, then $K = 1/\ln 2$ and in order to obtain one bit of information, there must be consumed an energy equal to $kT \ln 2$, which at $T = 300$ K is equivalent to 2×10^{-21} J. This is the lower estimate of the energy needed.

The amount of information contained in a living organism can be estimated roughly. According to Blumenfeld, the basic amount of information in the human organism is determined by the ordered arrangement of amino-acid residues in 7 kg of proteins, which corresponds to 3×10^{25} residues. This gives 1.3×10^{26} bits. The other contributions are considerably smaller: 150 g of DNA contained in the human organism corresponds to 6×10^{23} bits; the ordered arrangement of 10^{13} cells is represented by 4×10^{14} bits; and the ordered arrangement of 10^6 molecules of biopolymers in the cell is expressed by only 2.6×10^9 bits. The protein information is very small on the thermodynamic scale: 1.3×10^{26} k ln 2 = 1.3×10^3 J/K $\approx$ $\approx$ 300 cal/K. In entropy units the orderliness of a living organism is rather small; it is much smaller then the orderliness of a piece of rock of the same mass for the simple reason that the organism contains liquids.

The information theory and its applications in biology will be returned to in Sec. 17.9.

9.2. Nonequilibrium Processes

What has been said above refers to equilibrium systems in which the entropy is at a maximum. Living systems are nonequilibrium systems. We shall now turn to nonequilibrium systems and processes.

Let us consider a system consisting of two subsystems, say a gas whose volume is divided into two parts. We assume that the distributions of probabilities p_i' and p_i'' are specified in both subsystems. The corresponding entropies are given by

$$S' = -\mathrm{k} \sum_i p_i' \ln p_i' \quad \text{and} \quad S'' = -\mathrm{k} \sum_i p_i'' \ln p_i'' \tag{9.28}$$

As before, we have

$$\sum_i p_i' = 1 \quad \text{and} \quad \sum_i p_i'' = 1 \tag{9.29}$$

and

$$E' = \sum_i p_i' E_i' \quad \text{and} \quad E'' = \sum_i p_i'' E_i'' \tag{9.30}$$

The total energy is assumed to be constant:

$$E' + E'' = E \tag{9.31}$$

The entropy is additive:

$$S' + S'' = S \tag{9.32}$$

We have

$$\frac{\partial S}{\partial E'} = \frac{\partial S'}{\partial E'} + \frac{\partial S''}{\partial E'} = \frac{\partial S'}{\partial E'} + \frac{\partial S''}{\partial (E - E'')} = \frac{\partial S'}{\partial E'} - \frac{\partial S''}{\partial E''} \tag{9.33}$$

and, according to Eq. (9.20),

$$\frac{\partial S}{\partial E'} = \mathrm{k}\,(\beta' - \beta'') = \frac{1}{T'} - \frac{1}{T''} \tag{9.34}$$

In a nonequilibrium system all its characteristics, including the entropy, depend on time t. When the two subsystems, which have, say, different temperatures, are in contact with each other, a transfer of some physical quantity (say, energy) occurs between them. As a result of this transfer, the probability distribution is changed. We have

$$S(t) = -\mathrm{k} \sum_i p_i(t) \ln p_i(t) \tag{9.35}$$

Hence, in this particular case [see Eq. (9.34)],

$$\frac{dS}{dt} = \frac{\partial S}{\partial E'} \frac{dE'}{dt} = \left(\frac{1}{T'} - \frac{1}{T''} \right) \frac{dE'}{dt} \tag{9.36}$$

The change of the entropy with time is expressed by the product of the *generalized energy flux*

$$J = \frac{dE'}{dt} \tag{9.37}$$

by the *generalized force* defined as

$$X = \frac{1}{T'} - \frac{1}{T''} \tag{9.38}$$

In the above derivation we assumed that the entropy outside equilibrium depends on the same variables as it does in equilibrium—formula (9.35) is similar to formula (9.14). Without delving into subtle physical problems, we note that this assumption is proved by the nearness of the processes under consideration to equilibrium. A quantitative criterion of such nearness is given below (page 326).

If the system is characterized by many extensive variables (such as energy), the change of entropy with time is represented by the sum of the products of generalized forces and generalized fluxes:

$$\frac{dS}{dt} = \sum_j X_j J_j \tag{9.39}$$

The entropy change in an open system is made up of the entropy production inside the system, $d_i S$, and the entropy flow $d_e S$ (i.e., the exchange of entropy between the system and the outside world):

$$dS = d_i S + d_e S \tag{9.40}$$

According to the second law of thermodynamics, $d_i S \geqslant 0$; the sign of $d_e S$ has not been determined (see page 20). In the case under consideration, $d_e S = 0$ (the system as a whole is closed) and $dS = d_i S$—the entropy is produced due to the heat exchange between the two subsystems, as a result of which a thermal equilibrium will sooner or later be set up in the system. In a general case, expression (9.39) defines the so-called *dissipation function* σ—the rate of entropy production in unit volume. According to the second law of thermodynamics,

$$\frac{d_i S}{dt} = \int \sigma \, dV \geqslant 0 \tag{9.41}$$

and, according to Eq. (9.39),

$$\sigma = \sum_j J_j X_j \tag{9.42}$$

Here the forces and fluxes are referred to unit volume.

What generalized fluxes and generalized forces have we to deal with in the case of biological open systems—cells and organisms? We treat these systems as chemical systems that are at constant temperature. They undergo chemical reactions and transport of

matter occurs in them. For a chemical reaction the role of the flux is played by the reaction rate v, i.e., the derivative of the *reaction coordinate* ξ with respect to time:

$$J_{\text{chem}} = v = \frac{d\xi}{dt} \tag{9.43}$$

The coordinate ξ is the degree of advancement of the chemical reaction:

$$\xi(t) = \xi(0) + \int_0^t v\,dt = \xi(0) + \int_0^{\xi} \frac{dn_\gamma}{\nu_\gamma} \tag{9.44}$$

where n_γ is the number of moles of the reactant γ; ν_γ is the stoichiometric coefficient of this reactant in the reaction. The course of the chemical reaction is determined by the difference between the chemical potentials of the reactants and products, just as the heat energy flux is determined by the temperature difference. Accordingly, the generalized force for a chemical process is given by

$$X_{\text{chem}} = -\frac{1}{T}\sum_\gamma \nu_\gamma \mu_\gamma \equiv \frac{\mathcal{A}}{T} \tag{9.45}$$

where μ_γ is the chemical potential of component γ. The quantity $\mathcal{A} = -\sum_\gamma \nu_\gamma \mu_\gamma$ is called the *affinity* of the chemical reaction. The summation is carried out over all the reactants and products.

The chemical potential is given by

$$\mu_\gamma = \left(\frac{\partial G}{\partial n_\gamma}\right)_{n_{\gamma'},\ p,\ T} = \mu_\gamma^0(p,\ T) + RT \ln c_\gamma \tag{9.46}$$

where G is the Gibbs free energy; c_γ is the concentration (or activity) of the reactant γ.

The equilibrium constant of the reaction $K(p,\ T)$ is defined by the formula that expresses the law of mass action:

$$K(p,\ T) = c_{1\text{eq}}^{\nu_1} c_{2\text{eq}}^{\nu_2} \ldots c_{r,\text{eq}}^{\nu_r} = \exp\left(-\frac{1}{RT}\sum_{\gamma=1}^{r} \nu_\gamma \mu_\gamma^0\right) \tag{9.47}$$

The subscript "eq" signifies the equilibrium concentration. In a state of equilibrium

$$\mathcal{A}_{\text{eq}} = -\sum_\gamma \nu_\gamma \mu_\gamma^0 - RT\sum_\gamma \nu_\gamma \ln c_{\gamma\text{eq}} = \\ = -\sum_\gamma \nu_\gamma \mu_\gamma^0 - RT \ln K(p,\ T) = 0 \tag{9.48}$$

Hence, the affinity is defined by the formula

$$\mathcal{A} = RT \ln \frac{K(p, T)}{c_1^{\nu_1} c_2^{\nu_2} \dots c_r^{\nu_r}} \tag{9.49}$$

On the other hand, according to Eqs. (9.44), (9.45), and (9.46),

$$\mathcal{A} = -\sum_{\gamma} \nu_\gamma \frac{\partial G}{\partial n_\gamma}$$

But

$$\left(\frac{\partial G}{\partial \xi}\right)_{p,T} = \sum_{\gamma} \frac{\partial G}{\partial n_\gamma} \frac{dn_\gamma}{d\xi} = \sum_{\gamma} \frac{\partial G}{\partial n_\gamma} \nu_\gamma$$

Hence,

$$\mathcal{A} = -\left(\frac{\partial G}{\partial \xi}\right)_{p,T} \tag{9.50}$$

9.3. Coupling of Fluxes

The generalized fluxes J_i depend on the generalized forces and vice versa—the rate of a chemical reaction depends on the affinity, and the heat energy flux on the temperature difference. In a linear approximation

$$J_i = \sum_{j=1}^{n} L_{ij} X_j, \quad i, j = 1, 2, \dots, n \tag{9.51}$$

where L_{ij} are the *phenomenological coefficients.*

Simple examples of linear relationships are: the law of heat conduction (a heat flux, and the generalized force is the inverse-temperature difference); Ohm's law (the flux is the electric current and the force is the potential difference); Fick's law of diffusion (a flux of matter and the force is the concentration difference), etc.

Near the equilibrium, according to the *Onsager theorem*, the phenomenological coefficients L_{ij} form a symmetric matrix, i.e.,

$$L_{ij} = L_{ji} \tag{9.52}$$

This relation follows strictly from the principle of microscopic reversibility.

The relations inverse to Eq. (9.51) may be written in the form

$$X_i = \sum_{j=1}^{n} R_{ij} J_j \tag{9.53}$$

Here the coefficients R_{ij} also form a symmetric matrix.

Since $\sigma \geqslant 0$, certain constraints are imposed on the phenomenological coefficients L_{ij} (and, accordingly, on R_{ij}). Let us consider,

without loss of generality, the case of two forces and two fluxes:

$$\begin{aligned} J_1 &= L_{11}X_1 + L_{12}X_2 \\ J_2 &= L_{21}X_1 + L_{22}X_2 \end{aligned} \tag{9.54}$$

According to Eq. (9.42), we have

$$\sigma = L_{11}X_1^2 + (L_{12} + L_{21})\, X_1X_2 + L_{22}X_2^2 \geqslant 0 \tag{9.55}$$

The dissipation function σ is positive at any nonzero values of the variables X_1 and X_2 and vanishes if $X_1 = X_2 = 0$. Thus,

$$L_{11} > 0 \text{ and } L_{22} > 0 \tag{9.56}$$

and

$$(L_{12} + L_{21})^2 < 4L_{11}L_{22} \tag{9.57}$$

Since the coefficients are symmetrical, $L_{12} = L_{21}$,

$$L_{12}^2 < L_{11}L_{22} \tag{9.57a}$$

The sign of the off-diagonal coefficient L_{12} may be either positive or negative. In a general case,

$$L_{jj} > 0,\ L_{ii}L_{jj} > L_{ij} \tag{9.58}$$

It should be emphasized that the condition $\sigma \geqslant 0$ refers to the sum $\sum_i X_jJ_i$ as a whole. The individual members of this sum can be negative. This means that no single flux J_i is possible, since $X_iJ_i < 0$. In other words, such a flux would contradict the second law of thermodynamics. However, because of the coupling with other fluxes, for which we have positive values of $X_jJ_j > 0$, a flux inconceivable in a closed system becomes possible in an open system. The following condition must only be fulfilled:

$$\sum_{j \neq i} X_jJ_j > | X_iJ_i |$$

The coupling is determined by the difference of the off-diagonal coefficients L_{ij} from zero. Let us cite an example: a mixture of two gases in a vessel, whose walls are at different temperatures, spontaneously undergoes separation in such a manner that one gas predominates at the hot wall and the other at the cold wall. This phenomenon is called *thermal diffusion*. The matter flows in a direction opposite to the direction of the concentration gradient since the flow of matter is coupled with the flow of heat from the hot to the cold wall. The entropy deficiency in one process is compensated for by its excess production in the other.

We see that the entropy production in an open system makes possible the occurrence of processes that cannot take place in isolated systems. This is important for the understanding of biological systems.

The fluxes J_i and the forces X_i may be either scalar or vectorial. Suppose we have two fluxes—a scalar flux J_s and a vector flux $\boldsymbol{J}_v$:

$$J_s = L_{ss}X_s + \boldsymbol{L}_{sv}\boldsymbol{X}_v, \quad \boldsymbol{J}_v = \boldsymbol{L}_{vs}X_s + L_{vv}\boldsymbol{X}_v \qquad (9.59)$$

The coefficient L_{ss} is a scalar; $\boldsymbol{L}_{sv}$ and $\boldsymbol{L}_{vs}$ are vectors; finally, $\boldsymbol{L}_{vv}$ as the proportionality factor for the two vectors $\boldsymbol{X}_v$ and $\boldsymbol{J}_v$ is a tensor. If the system is isotropic, the force cannot produce a flux which has a different tensor dimension—a scalar cause cannot produce a vector effect and vice versa (the *Curie principle*). Hence, in this case, $\boldsymbol{L}_{sv} = \boldsymbol{L}_{vs} = 0$ and

$$J_s = L_{ss}X_s \quad \text{and} \quad \boldsymbol{J}_v = L_{vv}\boldsymbol{X}_v$$

No coupling between scalar and vector processes can occur in an isotropic system.

Let us prove this statement. A tensor quantity of rank n is changed upon orthogonal transformation of the coordinates as follows:

$$L'_{i_1,\, i_2 \ldots i_n} = \sum_{j_1 j_2 \ldots j_n} U(j_1, i_1)\, U(j_2,\, i_2) \ldots U(j_n,\, i_n)\, L_{j_1 j_2 \ldots j_n}$$

where j_k and i_k are Cartesian coordinates; $U(j_k, i_k)$ are the elements of the transformation matrix, whose determinant is equal to ± 1. If the system is isotropic, it is invariant with respect to the reflection in the centre (inversion), i.e., with respect to the transformation $x' = -x$, $y' = -y$, $z' = -z$, which may be written as follows:

$$r' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} r \equiv Ir$$

If $U = I$, the tensor L is transformed according to the law

$$\mathrm{L}' = (-1)^n \mathrm{L}$$

But because of the invariance $\mathrm{L}' = \mathrm{L}$ and, hence, all the coefficients at L with odd n vanish. In our case, for $\boldsymbol{L}_{sv}$ and $\boldsymbol{L}_{vs}$ we have $n = 1$. It means that these coefficients are equal to zero.

Because of the isotropic system being invariant with respect to rotations, the tensor L_{vv} assumes the form

$$\boldsymbol{L}_{vv} = L_{vv} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

that is, it becomes a scalar.

If the system is anisotropic but has a centre of symmetry (any crystals built up of symmetric molecules and devoid of screw symmetry), then the invariance with respect to inversion is preserved and, as before, $L_{sv} = L_{vs} = 0$.

9.4. Coupling of Chemical Reactions

Coupling of chemical reactions with one another and with diffusion is especially important in biology.

Let us consider a simplest chemical reaction near equilibrium:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B$$

The kinetic equation of the reaction has the form

$$\dot{c}_B = -\dot{c}_A = k_1 c_A - k_{-1} c_B \tag{9.60}$$

The flux, i.e., the reaction velocity, is given by

$$J_{\text{chem}} = v = -\dot{c}_A = \dot{c}_B \tag{9.61}$$

At equilibrium $v = 0$, $(\dot{c}_A)_{eq} = (\dot{c}_B)_{eq} = 0$, i.e.,

$$k_1 (c_A)_{eq} = k_{-1} (c_B)_{eq}$$

The equilibrium constant is given by

$$K = \frac{k_1}{k_{-1}} = \frac{(c_B)_{eq}}{(c_A)_{eq}}$$

Suppose that the system is a nonequilibrium system but is close to equilibrium. Then

$$c_A = (c_A)_{eq} + \alpha_A , \quad c_B = (c_B)_{eq} + \alpha_B$$

Here the deviations from equilibrium concentrations α_A and α_B are small:

$$\alpha_A \ll (c_A)_{eq}, \quad \alpha_B \ll (c_B)_{eq}$$

It is obvious that

$$c_A + c_B = (c_A)_{eq} + (c_B)_{eq}$$

and

$$\alpha_A + \alpha_B = 0$$

Under such conditions

$$J_{\text{chem}} = k_1 [(c_{Aeq} + \alpha_A)] - k_{-1} [(c_B)_{eq} + \alpha_B] =$$

$$= \alpha_A (k_1 + k_{-1}) = k_1 \alpha_A \frac{1+K}{K} \tag{9.62}$$

In this case, the affinity is equal to

$$\mathcal{A} = \mu_A - \mu_B \tag{9.63}$$

At equilibrium $(\mu_A)_{eq} = (\mu_B)_{eq}$ and $\mathcal{A} = 0$. Linear thermodynamics yields [cf. Eq. (9.45)]

$$J_{chem} = v = L\frac{\mathcal{A}}{T} = L\frac{\mu_A - \mu_B}{T} \tag{9.64}$$

where L is a phenomenological coefficient. Using the expression for the chemical potential, Eq. (9.46), we find

$$\begin{aligned}\mathcal{A} &= \mu_A^0 + RT\ln(c_A)_{eq} + RT\ln\left(1 + \frac{\alpha_A}{(c_A)_{eq}}\right) - \\ &- \mu_B^0 - RT\ln(c_B)_{eq} - RT\ln\left(1 + \frac{\alpha_B}{(c_B)_{eq}}\right) = \\ &= RT\left[\ln\left(1 + \frac{\alpha_A}{(c_A)_{eq}}\right) - \ln\left(1 + \frac{\alpha_B}{(c_B)_{eq}}\right)\right] \approx \\ &\approx RT\left[\frac{\alpha_A}{(c_A)_{eq}} - \frac{\alpha_B}{(c_B)_{eq}}\right] = RT\frac{\alpha_A}{(c_A)_{eq}}\frac{1+K}{K}\end{aligned} \tag{9.65}$$

and, consequently,

$$J_{chem} = \frac{LR}{(c_A)_{eq}}\alpha_A\frac{1+K}{K} \tag{9.66}$$

Comparing Eq. (9.66) with Eq. (9.62), we find the value of the coefficient L:

$$L = \frac{k_1(c_A)_{eq}}{R} \tag{9.67}$$

Since $|\alpha_A| \ll (c_A)_{eq}$ and $|\alpha_B|/(c_B)_{eq} = |\alpha_A|/(c_B)_{eq} \ll 1$, we get:

$$\frac{\alpha_A}{(c_A)_{eq}} \ll \frac{(c_B)_{eq}}{(c_A)_{eq}} = K$$

Thus,

$$\frac{\alpha_A}{(c_A)_{eq}} \ll K\left(1 - \frac{\alpha_A}{(c_A)_{eq}}\right) \approx K$$

or

$$\frac{\alpha_A}{(c_A)_{eq}}\frac{1+K}{K} \ll 1$$

that is, according to Eq. (9.65),

$$\frac{\mathcal{A}}{RT} \ll 1 \tag{9.68}$$

This is the near-equilibrium condition, which is equivalent to $\alpha_A \ll (c_A)_{eq}$. We shall prove that this condition directly yields the linear relation (9.64). Suppose that the reaction velocity is de-

fined as follows:

$$v = \dot{c}_B = k_1 c_A - k_{-1} c_B = k_1 c_A \left(1 - \frac{k_{-1}}{k_1}\frac{c_B}{c_A}\right) = k_1 c_A \left(1 - \frac{1}{K}\frac{c_B}{c_A}\right)$$

and the affinity $\mathcal{A}$ is given by

$$\mathcal{A} = RT \ln \frac{K}{c_A^{-1} c_B}$$

From this it follows that

$$\frac{c_B}{c_A} = K \exp\left(-\frac{\mathcal{A}}{RT}\right)$$

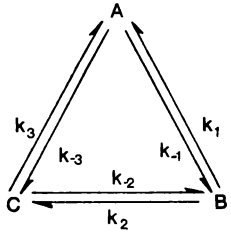
However, if the near-equilibrium condition (9.68) is valid, then, expanding the exponential into a series and retaining only the linear term, we get:

$$v = \frac{k_1 c_A}{R}\frac{\mathcal{A}}{T} = L\frac{\mathcal{A}}{T}$$

Near equilibrium $c_A \approx (c_A)_{eq}$. Hence, we again have [see Eq. (9.67)]

$$L = \frac{k_1 (c_A)_{eq}}{R}$$

Let us now consider the coupling of several reactions. Suppose that three reactions are involved in the *cycle*



The fluxes are given by

$$\begin{aligned} J_1 &= k_1 c_A - k_{-1} c_B \\ J_2 &= k_2 c_B - k_{-2} c_C \\ J_3 &= k_3 c_C - k_{-3} c_A \end{aligned} \tag{9.69}$$

The affinities are equal to

$$\begin{aligned} &\mathcal{A}_1 = \mu_A - \mu_B, \quad \mathcal{A}_2 = \mu_B - \mu_C \\ &\mathcal{A}_3 = \mu_C - \mu_A, \quad \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 = 0 \end{aligned} \tag{9.69a}$$

The dissipation function expressed through the independent variables and multiplied by T is given by

$$\sigma T = J_1 \mathcal{A}_1 + J_2 \mathcal{A}_2 + J_3 \mathcal{A}_3 = (J_1 - J_3)\,\mathcal{A}_1 + (J_2 - J_3)\,\mathcal{A}_2 \tag{9.70}$$

The fluxes $J_1 - J_3$ and $J_2 - J_3$ are independent. We have

$$\begin{aligned} J_1 - J_3 &= L_{11}\mathcal{A}_1 + L_{12}\mathcal{A}_2 \\ J_2 - J_3 &= L_{21}\mathcal{A}_1 + L_{22}\mathcal{A}_2 \end{aligned} \tag{9.71}$$

At equilibrium $\mu_A^{eq} = \mu_B^{eq} = \mu_C^{eq}$, $\mathcal{A}_1^{eq} = \mathcal{A}_2^{eq} = 0$, $J_1 - J_3 = J_2 - J_3 = 0$, and $J_1 = J_2 = J_3 = 0$. The fluxes are equal to zero according to the principle of detailed balance, which yields an important relation for cyclic kinetic processes:

$$k_1 k_2 k_3 = k_{-1} k_{-2} k_{-3} \tag{9.72}$$

Near equilibrium

$$\begin{aligned} J_1 &= k_1 \alpha_A - k_{-1} \alpha_B \\ J_2 &= k_2 \alpha_B - k_{-2} \alpha_C \\ J_3 &= k_3 \alpha_C - k_{-3} \alpha_A \end{aligned} \tag{9.73}$$

and

$$\begin{aligned} \mathcal{A}_1 &= RT \left(\frac{\alpha_A}{(c_A)_{eq}} - \frac{\alpha_B}{(c_B)_{eq}} \right) = \frac{RT}{k_1 (c_A)_{eq}} (k_1 \alpha_A - k_{-1} \alpha_B) \\ \mathcal{A}_2 &= RT \left(\frac{\alpha_B}{(c_B)_{eq}} - \frac{\alpha_C}{(c_C)_{eq}} \right) = \frac{RT}{k_2 (c_B)_{eq}} (k_2 \alpha_B - k_{-2} \alpha_C) \end{aligned} \tag{9.74}$$

from which

$$J_1 = \frac{k_1 (c_A)_{eq}}{RT} \mathcal{A}_1, \quad J_2 = \frac{k_2 (c_B)_{eq}}{RT} \mathcal{A}_2, \quad J_3 = -\frac{k_3 (c_C)_{eq}}{RT} (\mathcal{A}_1 + \mathcal{A}_2)$$

Hence,

$$\begin{aligned} J_1 - J_3 &= \frac{k_1 (c_A)_{eq} + k_3 (c_C)_{eq}}{R} \frac{\mathcal{A}_1}{T} + \frac{k_3 (c_C)_{eq}}{R} \frac{\mathcal{A}_2}{T} \\ J_2 - J_3 &= \frac{k_3 (c_C)_{eq}}{R} \frac{\mathcal{A}_1}{T} + \frac{k_2 (c_B)_{eq} + k_3 (c_C)_{eq}}{R} \frac{\mathcal{A}_2}{T} \end{aligned} \tag{9.75}$$

So we find that

$$L_{11} = \frac{k_1 (c_A)_{eq} + k_3 (c_C)_{eq}}{R} \qquad L_{22} = \frac{k_2 (c_B)_{eq} + k_3 (c_C)_{eq}}{R}$$

and

$$L_{12} = L_{21} = \frac{k_3 (c_C)_{eq}}{R}$$

The Onsager theorem $L_{ij} = L_{ji}$ is automatically fulfilled in the neighbourhood of equilibrium.

Coupling of chemical reactions in an open system makes possible the occurrence of endergonic reactions forbidden in closed systems, because in these reactions the free energy is increased. We have already encountered such processes, in particular those involved in the formation of protein chains. The formation of each peptide bond

occurs with release of one water molecule. Since the cell contains an excess amount of water, the inverse reaction of hydrolysis of peptide bonds must prevail. But, as we have seen, the polycondensation is coupled to the exergonic reaction of cleavage of ATP (see Chapter 8) and the dissipation function is positive on the whole. The coupling of endergonic processes to the hydrolysis of ATP is of general importance in biology. Coupling makes possible for ATP to perform its universal role as a source of free energy required for endergonic processes to occur. If cells and organisms were isolated systems, ATP would not be able to play this role.

Thus, nonequilibrium thermodynamics even in its linear approximation proves the possibility of occurrence, in open systems, of processes forbidden in closed systems. This is of paramount importance in biology.

9.5. Steady States of Linear Systems

As has already been said (page 20), an open system may be in a steady, though nonequilibrium, state. In this state, the entropy production inside the system is exactly compensated for by the flow of entropy to the outside world, so that the total entropy of the system is not changed (see page 20):

$$dS = d_i S + d_e S = 0 \tag{9.76}$$

Bertalanffy, one of the founders of the thermodynamics of open biological systems, termed such a stationary state the state of flow equilibrium. It differs from true equilibrium by many specific features.

The stationary state of an open system is achieved if the system is subject to constraints that fix the constant values of a certain set of generalized forces, but the other generalized forces may vary. If X_1, X_2, . . ., X_j are constant and X_{j+1}, X_{j+2}, . . ., X_n freely change, then

$$J_i = 0 \text{ for } i > j \quad \text{and} \quad J_i \neq 0 \text{ for } i \leqslant j$$

The entropy of such a system does not attain a maximum and the dissipation function σ is different from zero. Suppose that the system exchanges matter and energy with the environment but its mass remains unchanged. There is a flux of heat J_T in the system and the flux of matter $J_M = 0$. Hence,

$$\sigma = J_T X_T + J_M X_M \tag{9.77}$$

and

$$\begin{aligned} J_T &= L_{11} X_T + L_{12} X_M \neq 0 \\ J_M &= L_{21} X_T + L_{22} X_M = 0 \end{aligned} \tag{9.78}$$

Assuming the force X_T (say, a temperature difference) to be fixed, we have:

$$\sigma = L_{11}X_T^2 + (L_{12} + L_{21})\, X_T X_M + L_{22}X_M^2 \tag{9.79}$$

Differentiating this expression with respect to X_M at X_T = const yields:

$$\frac{\partial \sigma}{\partial X_M} = 2L_{22}X_M + (L_{12} + L_{21})\, X_T$$

If the system is close to equilibrium and the Onsager theorem is fulfilled, then $L_{12} = L_{21}$ and

$$\frac{\partial \sigma}{\partial X_M} = 2\,(L_{12}X_T + L_{22}X_M) = 2J_M = 0 \tag{9.80}$$

The second derivative

$$\frac{\partial^2 \sigma}{\partial X_M^2} = 2L_{22} > 0$$

Thus, in the stationary state close to equilibrium the entropy production σ is a minimum. This is the *Prigogine theorem*. It is directly associated with Onsager's theorem and is valid only within the limits of applicability of linear thermodynamics, i.e., when the coefficients L_{ij} are independent of X_i.

If X_T can also be changed in an arbitrary way, then $J_T = J_M = 0$, $\sigma = 0$ and the system reaches equilibrium.

Let us consider an open chemical system. Substance A is introduced into it from the outside world and undergoes a series of consecutive conversions inside the system, and the final product F leaves the system:

$$\longrightarrow \mathrm{A} \xrightarrow{1} \mathrm{B} \xrightarrow{2} \mathrm{C} \longrightarrow \ldots \xrightarrow{r} \mathrm{F} \longrightarrow$$

The kinetic equations are as follows:

$$\frac{dn_A}{dt} = \frac{d_{ex}n_A}{dt} - v_1, \quad \frac{dn_B}{dt} = v_1 - v_2, \ldots, \quad \frac{dn_F}{dt} = \frac{d_{ex}n_F}{dt} + v_r \tag{9.81}$$

where the derivatives with the subscript "ex" characterize the exchange with the environment.

In the stationary state the content of all the reactants is constant in a flow-through system:

$$\dot{n}_i = 0 \qquad (i = \mathrm{A}, \mathrm{B}, \ldots \mathrm{F})$$

Hence, all the reaction velocities are equilibrated:

$$\frac{d_{ex}n_A}{dt} = v_1 = v_2 = \ldots = v_r = -\frac{d_{ex}n_F}{dt} = v \tag{9.82}$$

This may be illustrated by a spectacular model. The progress of a chemical reaction may be visualized as the flow of a liquid. If the system is a closed one, i.e., if the liquid neither flows into nor emerges from the vessels, then the entire liquid will flow at a velocity determined by the tap aperture from the upper vessel to the lower one

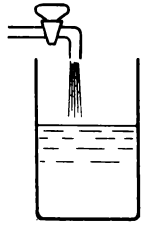


Fig. 9.1. The model of a closed system.

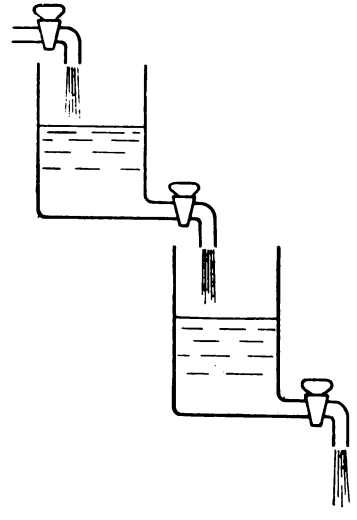


Fig. 9.2. The model of an open system.

and a state of equilibrium will set in (Fig. 9.1). The level of the liquid in the lower vessel will represent the extent of reaction in such a state.

If the system is an open, flow-through system, a certain level of the liquid will be set up in the upper and in the lower vessel which will not correspond to equilibrium (Fig. 9.2). This level will depend on the steady-state flow rate, i.e., on the extent to which the tap is opened. We see that the tap models the catalyst. In a closed system the final extent of reaction does not depend on its rate—on the extent to which the tap is opened; there will eventually be set up a constant level of the liquid in the lower vessel. In an open system, not only the rate but the extent of reaction as well depends on the catalyst. When the extent of opening of the tap is changed, new liquid levels, a new steady state will be attained.

Let us prove that the condition of constant reaction rates is equivalent to minimum entropy production. For our system we have

$$T\sigma = \mathcal{A}_A \frac{d_{ex}n_A}{dt} + \sum_{i=1}^{r} \mathcal{A}_i v_i + \mathcal{A}_F \frac{d_{ex}n_F}{dt} \tag{9.83}$$

where

$$\mathcal{A}_A = \mu_A^{(1)} - \mu_A^{(2)}, \quad \mathcal{A}_F = \mu_F^{(1)} - \mu_F^{(2)}$$

where the superscripts 1 and 2 refer to the system and the environment, respectively. In the stationary state

$$T\sigma = \left(\mathcal{A}_A + \sum_i \mathcal{A}_i - \mathcal{A}_F\right) v = \mathcal{A}v > 0 \tag{9.84}$$

where $\mathcal{A}$ corresponds to the overall process $A^{(2)} \rightarrow F^{(2)}$. On the other hand, in the linear range,

$$\sigma = \sum_{i=1}^{r+2} \sum_{j=1}^{r+2} L_{ij} \frac{\mathcal{A}_i}{T} \frac{\mathcal{A}_j}{T} > 0 \tag{9.85}$$

The minimum in Eq. (9.85) is determined at a given value of total affinity $\mathcal{A}$ with the aid of the Lagrange multiplier λ. We differentiate the function

$$\Phi = \sum_i \sum_j L_{ij} \frac{\mathcal{A}_i}{T} \frac{\mathcal{A}_j}{T} - 2\lambda \sum_i \frac{\mathcal{A}_i}{T}$$

with respect to $\mathcal{A}_i/T$ and set the derivative to zero. We thus obtain

$$\frac{\partial \Phi}{\partial (\mathcal{A}_i/T)} = 2 \sum_j L_{ij} \frac{\mathcal{A}_j}{T} - 2\lambda = 0$$

The reaction velocity

$$v_i = \sum_j L_{ij} \mathcal{A}_j / T = \lambda = \text{const}$$

Hence, the sought-for equivalency is proved.

It has already been pointed out that for the steady state to be attained, it is necessary that two time scales be available. The state of the biosphere is approximately stationary during the time much shorter than the time of cooling of the Sun. Steady-state fluxes of matter and energy (*homeostasis* or, exactly, *homeorhesis* according to Waddington) are maintained in the organism during the time period much shorter than the time elapsed from birth to death.

Let us now determine the stability conditions for the stationary state. The Prigogine theorem of minimum entropy production gives a criterion of evolution, showing that the system necessarily passes over to a steady nonequilibrium state from any state close to it.

Without loss of generality we shall consider two coupled chemical reactions. The dissipation function is positive:

$$\sigma_{in} = L_{11}\left(\frac{\mathcal{A}_1}{T}\right)^2 + 2L_{12}\frac{\mathcal{A}_1}{T}\frac{\mathcal{A}_2}{T} + L_{22}\left(\frac{\mathcal{A}_2}{T}\right)^2 > 0 \tag{9.86}$$

We calculate $\dot{\sigma}_{in}$, assuming the coefficients to be constant:

$$\frac{1}{2}\dot{\sigma}_{in} = \left(L_{11}\frac{\mathcal{A}_1}{T} + L_{12}\frac{\mathcal{A}_2}{T}\right)\frac{d(\mathcal{A}_1/T)}{dt} + \left(L_{12}\frac{\mathcal{A}_1}{T} + L_{22}\frac{\mathcal{A}_2}{T}\right) \times$$
$$\times \frac{d(\mathcal{A}_2/T)}{dt} = v_1\frac{d(\mathcal{A}_1/T)}{dt} + v_2\frac{d(\mathcal{A}_2/T)}{dt} \qquad (9.87)$$

In a closed system $\mathcal{A}_1$ and $\mathcal{A}_2$ depend on the reaction coordinates ξ_1 and ξ_2 and also on pressure and temperature, which are assumed here to be constant. We have

$$\frac{1}{2}\dot{\sigma}_{in} = \frac{v_1}{T}\left(\frac{\partial\mathcal{A}_1}{\partial\xi_1}v_1 + \frac{\partial\mathcal{A}_1}{\partial\xi_2}v_2\right) + \frac{v_2}{T}\left(\frac{\partial\mathcal{A}_2}{\partial\xi_1}v_1 + \frac{\partial\mathcal{A}_2}{\partial\xi_2}v_2\right) \qquad (9.88)$$

We have seen that the affinity $\mathcal{A}$ is the derivative of the free energy of the system with respect to the coordinate of the corresponding reaction taken with a minus sign, i.e., the equality (9.50) is valid:

$$\mathcal{A}_i = -\frac{\partial G}{\partial\xi_i}$$

Thus, $\partial\mathcal{A}_1/d\xi_2 = \partial\mathcal{A}_2/\partial\xi_1$ and we have

$$\frac{1}{2}\dot{\sigma}_{in} = \frac{1}{T}\left(\frac{\partial\mathcal{A}_1}{\partial\xi_1}v_1^2 + 2\frac{\partial\mathcal{A}_1}{\partial\xi_2}v_1v_2 + \frac{\partial\mathcal{A}_2}{\partial\xi_2}v_2^2\right) < 0 \qquad (9.89)$$

The quantity $\dot{\sigma}_{in}$ is negative, since the values of $\mathcal{A}_i$ decrease with increasing ξ_i: at equilibrium $\mathcal{A}_i = 0$.

In a closed system the entropy production can only decrease with time. In an open system there is an exchange of matter with the environment, and the overall entropy change is described by the following expression:

$$\frac{1}{2}\dot{\sigma} = \frac{1}{2}\dot{\sigma}_{in} + \frac{1}{T}\sum_{\gamma}\left(v_1\frac{\partial\mathcal{A}_1}{\partial n_\gamma} + v_2\frac{\partial\mathcal{A}_2}{\partial n_\gamma}\right)\frac{d_{ex}n_\gamma}{dt} \qquad (9.90)$$

Here the first term is always positive and the sign of the second depends on the concrete situation. The inner irreversible processes invariably lower the rate of entropy production. From the inequality $\dot{\sigma}_{in} < 0$ it follows that if the system has reached a state with a minimum of entropy production, it cannot leave it spontaneously.

The differential of the dissipation function may be written in the following form:

$$d\sigma = d_J\sigma + d_X\sigma = \sum_i X_i\,dJ_i + \sum_i J_i\,dX_i \qquad (9.91)$$

In the linear range, by virtue of Onsager's theorem, $d_X\sigma = d_J\sigma = {}^1/_2 d\sigma$, i.e., $d_X\sigma$ is the exact differential (it is precisely this differential that appears in Eq. (9.88), in which the fluxes v_i are assumed

to be constant). Hence,

$$\frac{d_X\sigma}{dt} = \frac{1}{2}\frac{d\sigma}{dt}$$

In the stationary state $\dot{\sigma}_{in} < 0$, i.e., $d_X\sigma < 0$, $d\sigma < 0$. Accordingly, the variation condition of stability consists in that

$$\delta\sigma = 2\delta_X\sigma \geqslant 0 \tag{9.92}$$

The sign of the variation is opposite to the sign of the differential—a fluctuation can only give rise to excess entropy production. Let us assume that $J_M = 0$ in the stationary state (see page 329). At fluctuations δX_M we have $J_M = L_{MM}\delta X_M$, where $L_{MM} > 0$ and

$$\delta_X\sigma = J_M\delta X_M = L_{MM}(\delta X_M)^2 > 0$$

For a chemical reaction this condition has the form

$$T\delta_X\sigma = \sum_i v_i\delta\mathcal{A}_i > 0$$

In the stationary state

$$Td_X\sigma = \sum_i v_i^{st} d\mathcal{A}_i = 0$$

if all $\mathcal{A}_i$'s are independent. This can always be attained through obtaining appropriate linear combinations of reaction velocities and affinities. Hence, we may put the steady-state values of v_i^{st} equal to zero. It means that

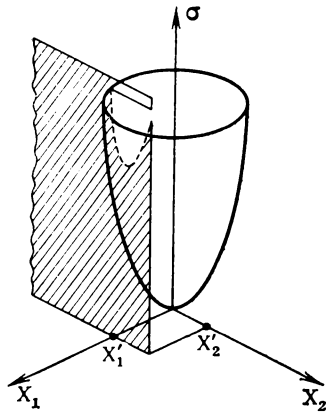


Fig. 9.3. A diagram illustrating the behaviour of a steady-state linear system.

$$v_i = v_i^{st} + \delta v_i = \delta v_i$$

and the stability condition for the stationary state will have the following form:

$$T\delta_X\sigma = \sum_i \delta v_i\,\delta\mathcal{A}_i > 0 \tag{9.93}$$

As we have already seen [see, for example, Eq. (9.79)], the quantity σ is expressed by a quadratic function of X. This is depicted schematically in Fig. 9.3 for the case of two forces, X_1 and X_2. If at $X_1 = \text{const}$ the force X_2 can change, the point representing the quantity σ will move along the parabola obtained upon intersection of the

surface σ (X_1, X_2) with the $X_1 = \text{const}$ plane until it attains a minimum of σ. If there are no constraints, the quantity σ will move across the surface until it reaches the equilibrium point $\sigma = 0$.

The return of the deviated linear system to a stationary state close to equilibrium occurs in an exponential manner, without oscillations:

$$\delta\xi = \xi(t) - \xi^{st} = (\xi(0) - \xi^{st}) \exp(-t/\tau) \qquad (9.94)$$

where ξ^{st} is the steady-state value of the reaction coordinate and τ is the *relaxation time.*

9.6. Coupling of Chemical Reactions to Transport of Matter

The organism and the cell are chemical machines that operate as a result of chemical reactions and transport of matter between the cell and the surrounding medium and also inside the cell. This transport occurs in a certain definite direction perpendicular to the cellular and intracellular membranes. The flux of matter is a vector. At the same time, the rate of a chemical reaction is a scalar. As has been pointed out (page 324), the direct coupling of scalar and vector processes is impossible in an isotropic system by virtue of the Curie principle. Nor is it possible in anisotropic systems having a centre of symmetry. Biological systems, in which chemical reactions are coupled to diffusion (namely membranes) are, however, made up of chiral molecules which have neither a plane nor a centre of symmetry (Sec. 2.7). Biological membranes are anisotropic. In such systems, direct coupling is, in principle, possible; the vectorial coefficients $L_{sv} = L_{vs}$ may be different from zero. No theory of direct coupling between chemical reactions and diffusion in membranes has been worked out so far. One can visualize, for example, the transport of some reaction participants along the screw channel in the membrane in which the centres are located. Then, the course of the reaction will be different for substances supplied from the opposite ends of the channel. The same result is obtained by considering a symmetrical channel in which asymmetric (i.e., chiral) reaction centres are regularly arranged. There is, however, no ground so far for stating that these effects are considerable.

The interplay between chemical reactions and the transport of matter can also take place by way of *indirect coupling* which arises due to the steady-state condition. In the stationary state, the irreversible processes that are not directly coupled become interlinked. Let us return to the example considered on page 330. Along with substances that participate in the reaction

$$\rightarrow A \rightarrow \ldots \rightarrow F \rightarrow$$

we shall introduce into the system an inert substance Q which will not take part in the reaction. If the transport of this substance is,

however, coupled to the transport of A, then

$$\begin{aligned} T\frac{d_{ex}n_A}{dt} &= L_{21}\mathcal{A}_A + L_{12}\mathcal{A}_Q \\ T\frac{d_{ex}n_Q}{dt} &= L_{12}\mathcal{A}_A + L_{22}\mathcal{A}_Q \\ T\frac{d_{ex}n_F}{dt} &= L_F\mathcal{A}_F \\ Tv &= L_{chem}\mathcal{A}_{chem} \end{aligned} \tag{9.95}$$

The last equation refers to the overall chemical reaction A $\rightarrow$ F. From the steady-state conditions it follows that

$$\frac{dn_A}{dt} = \frac{d_{ex}n_A}{dt} - v = 0, \quad \frac{dn_F}{dt} = \frac{n_{ex}n_F}{dt} + v = 0,$$

$$\frac{dn_Q}{dt} = \frac{d_{ex}n_Q}{dt} = 0$$

We obtain

$$\mathcal{A}_A = \frac{Tv}{L_{11} - L_{12}^2/L_{22}}, \quad \mathcal{A}_F = -\frac{Tv}{L_F}, \quad \mathcal{A}_Q = -\frac{TvL_{12}/L_{22}}{L_{11} - L_{12}^2/L_{22}} \tag{9.96}$$

The affinity $\mathcal{A}_Q$ of the inert component is different from zero and proportional to the rate of the chemical reaction in which substance Q does not participate. This reaction gives rise to a difference in concentrations of substance Q inside and outside the system:

$$c_Q^{(1)}/c_Q^{(2)} = K \exp(-\mathcal{A}_Q/RT) \tag{9.97}$$

This provides an indirect coupling of the transport of substance Q to the chemical reaction.

Let us consider indirect coupling in the general form (Katchalsky and Oster). We shall write the dissipation function for a system in which a transport (the generalized force is the chemical potential gradient) and a chemical reaction take place:

$$T\sigma = \sum_i J_i \nabla(-\mu_i) - v\sum_i \nu_i \mu_i \tag{9.98}$$

The continuity condition for the ith component is

$$\frac{\partial c_i}{\partial t} = -\nabla J_i + \nu_i v \tag{9.99}$$

The steady-state condition, i.e., the constancy of the concentration c_i, is $\dot{c}_i = 0$, i.e.,

$$\nabla J_i = \nu_i v \tag{9.100}$$

Substituting Eq. (9.100) into Eq. (9.98), we find that

$$T\sigma = -\nabla \sum_i J_i \mu_i \tag{9.101}$$

We integrate σ for a one-dimensional flow along the x axis from $x = 0$ to $x = a$. The total dissipation per unit surface area is equal to

$$\sigma_{\text{total}} = \int_0^a \sigma \, dx = \frac{1}{T} \sum_i [J_i(0)\,\mu_i(0) - J_i(a)\,\mu_i(a)] \tag{9.102}$$

An analogous integration of ∇J_i yields [cf. Eq. (9.100)]:

$$\int_0^a \frac{dJ_i}{dx}\, dx = J_i(a) - J_i(0) = \nu_i \int_0^a v\, dx = \nu_i J_{\text{chem}}^{\text{total}} \tag{9.103}$$

From the two last equations it follows that

$$T\sigma_{\text{total}} = \sum_i J_i(0)\,[\mu_i(0) - \mu_i(a)] + J_{\text{chem}}^{\text{total}} \sum_i [-\nu_i \mu_i(a)] \tag{9.104}$$

or

$$T\sigma_{\text{total}} = \sum_i J_i(a)\,[\mu_i(0) - \mu_i(a)] + J_{\text{chem}}^{\text{total}} \sum_i [-\nu_i \mu_i(0)] \tag{9.104a}$$

The sums in the last terms on the right-hand sides of these expressions represent the values of reaction affinities at the membrane boundaries $x = 0$ and $x = a$ for the components that penetrate into the membrane. Passing to the phenomenological equations, we write:

$$\begin{aligned} J_i(0) &= \sum_j L_{ij}(0)\,\Delta\mu_j + L_{i,\,\text{chem}}(0)\,\mathcal{A}(a) \\ J_{\text{chem}}^{\text{total}} &= \sum_i L_{\text{chem},\,i}(0)\,\Delta\mu_i + L_{\text{chem}}(0)\,\mathcal{A}(a) \end{aligned} \tag{9.105}$$

or

$$\begin{aligned} J_i(a) &= \sum_j L_{ij}(a)\,\Delta\mu_j + L_{i,\,\text{chem}}(a)\,\mathcal{A}(0) \\ J_{\text{chem}}^{\text{total}} &= \sum_i L_{\text{chem},\,i}(a)\,\Delta\mu_i + L_{\text{chem}}(a)\,\mathcal{A}(0) \end{aligned} \tag{9.105a}$$

The macroscopic, i.e., integrated, expressions now contain off-diagonal coefficients $L_{i,\text{chem}}(0)$ and $L_{i,\text{chem}}(a)$. But are they symmetric? The local, unintegrated expressions do not contain off-diagonal coefficients for chemical reactions and diffusion, since the system is isotropic and the Curie principle is valid. To put it differently,

$$J_i = \sum_j l_{ij} \nabla(-\mu_j), \quad v = l_{\text{chem}} \mathcal{A} \tag{9.106}$$

where l's are local phenomenological coefficients. The same refers to the inverse expressions:

$$\nabla(-\mu_j) = \sum_i r_{ji} J_i, \quad \mathcal{A} = r_{\text{chem}} v = l_{\text{chem}}^{-1} v \tag{9.107}$$

Let us find the relation between the local and macroscopic coefficients. Applying the operator ∇ to the first equation in (9.107), we obtain

$$\nabla^2(-\mu_i) = \sum_j r_{ij} \nabla J_j$$

Substituting the steady-state condition (9.100), we find that

$$\nabla^2(-\mu_i) = \left(\sum_j \nu_j r_{ij}\right) v = \left(l_{\text{chem}} \sum_j \nu_j r_{ij}\right) \mathcal{A}$$

We multiply both sides of this equation by ν_i and carry out the summation over i. Since $\mathcal{A} = -\sum_i \nu_i \mu_i$, we obtain

$$\sum_i \nu_i \nabla^2(-\mu_i) = \nabla^2 \mathcal{A} = \left(l_{\text{chem}} \sum_{i,j} \nu_i \nu_j r_{ij}\right) \mathcal{A} \qquad (9.108)$$

We have obtained a differential equation for $\mathcal{A}$. The coefficient at $\mathcal{A}$ has the dimensions of cm^{-2}. We designate it by λ^{-2} (λ is the relaxation length). We have

$$\lambda^2 \nabla^2 \mathcal{A} = \mathcal{A} \qquad (9.108a)$$

The solution of this equation has the form

$$\mathcal{A}(x) = \frac{\mathcal{A}(a) \sinh(x/\lambda) + \mathcal{A}(0) \sinh[(a-x)/\lambda]}{\sinh(a/\lambda)} \qquad (9.109)$$

In order to pass over to macroscopic expressions, we have to integrate the local relations. Integration of Eq. (9.100) from 0 to x gives

$$\int_0^x \frac{dJ_i}{dx}\, dx = J_i(x) - J_i(0) = \nu_i l_{\text{chem}} \int_0^x \mathcal{A}\, dx \qquad (9.110)$$

Substituting Eq. (9.109) and integrating, we arrive at the following expression:

$$J_i(x) = J_i(0) + \frac{\nu_i \lambda l_{\text{chem}}}{\sinh(a/\lambda)} \{\mathcal{A}(0)(\cosh(x/\lambda) - 1) - \\ - \mathcal{A}(0)(\cosh[(a-x)/\lambda] - \cosh(a/\lambda))\} \qquad (9.111)$$

On the other hand, according to Eq. (9.106), the left-hand side of this expression is equal to

$$J_i(x) = \sum_j l_{ij} \nabla(-\mu_j)$$

We substitute this expression into Eq. (9.111) and integrate from 0 to a. Since

$$-\int_0^a \frac{d\mu_i}{dx}\, dx = \mu_i(0) - \mu_i(a) = \Delta\mu_i$$

and

$$\mathcal{A}(a) - \mathcal{A}(0) = -\sum_i \nu_i \mu_i(a) + \sum_i \nu_i \mu_i(0) = \sum_i \nu_i \Delta\mu_i$$

then, through integration, we obtain

$$J_i(0) \cdot a = \sum_j [l_{ij} + \nu_i \nu_j \lambda l_{\text{chem}} (a \coth(a/\lambda) - \lambda)] \Delta\mu_j - \\ - [\nu_i \lambda l_{\text{chem}} a \tanh(a/2\lambda)] \mathcal{A}(a) \quad (9.112)$$

We divide expression (9.112) by a and obtain macroscopic phenomenological expressions. We introduce the notation

$$\lambda l_{\text{chem}} [\coth(a/\lambda) - \lambda/a] = \alpha, \quad \lambda l_{\text{chem}} \tanh(a/2\lambda) = \beta$$

Finally we find that

$$J_i(0) = \sum_j (l_{ij}/a + \nu_i \nu_j \alpha) \Delta\mu_j - \nu_i \beta \mathcal{A}(a) \quad (9.113)$$

which expression coincides with Eq. (9.105), with

$$L_{ij}(0) = l_{ij}/a + \nu_i \nu_j \alpha, \quad L_{i,\text{chem}}(0) = -\nu_i \beta$$

Calculation of $J_{\text{chem}}^{\text{total}}$ gives

$$J_{\text{chem}}^{\text{total}} = \int_0^a J_{\text{chem}}\, dx = l_{\text{chem}} \int_0^a \mathcal{A}\, dx = \lambda l_{\text{chem}} [\mathcal{A}(a) + \\ + \mathcal{A}(0)] \tanh(a/2\lambda) = -\sum_i \nu_i \beta \Delta\mu_i + 2\beta \mathcal{A}(a) \quad (9.114)$$

Comparing Eq. (9.114) with the second equation in (9.105), we find:

$$L_{\text{chem},i}(0) = -\nu_i \beta, \quad L_{\text{chem}}(0) = 2\beta \quad (9.115)$$

Thus, the macroscopic coefficients, which characterize the indirect coupling of diffusion fluxes to a chemical reaction, are expressed through microscopic coefficients. The Onsager reciprocity relations remain to be valid. That the macroscopic coefficients of coupling are different from zero is determined in an isotropic system by the stationary state conditions.

As we shall see later, indirect coupling plays an important role in the theory of membrane transport (see Chapters 10 and 13). In Chapter 10 we shall discuss indirect coupling for the so-called *facilitated transport* which involves molecules of *carriers*.

9.7. Far-from-Equilibrium Processes

Living systems are characterized by a high structural order and ordered behaviour in space and time. We have already pointed out the seeming contradiction between the increasing complexity of the

system in the course of its biological development and the second law of thermodynamics (page 14). This contradiction is easily removed: a living system is an open system whose entropy may increase or fall off. Usually, one speaks of the "anti-entropy nature" of life. However, these words can in no way explain the specific features of the space-time order of an open system, and the meaning of the "anti-entropy" concept remains to be obscure.

In nature we encounter two types of order—static order and dynamic order. In the first case, order is attained under thermodynamic equilibrium conditions when the temperature is sufficiently decreased, say, upon crystallization of a liquid. Static order arises as a result of a phase transition, which occurs under equilibrium conditions. This kind of order is not practically found in biology—the Schrödinger *aperiodic crystal* (page 14) is fundamentally different from an equilibrium periodic crystal. The dynamic order of a living system is realized not because the entropy is diminished due to a decrease in temperature; it occurs because there is an outflow of entropy from the open system into the surroundings. The appearance of a space-time structure is in this case too of the nature of a *phase transition*, but a nonequilibrium one this time. Investigations of dynamic order, which are of prime importance in physics and biology, have been started relatively recently. A number of model nonbiological systems (in particular, chemical systems) are known at present, in which there is observed a dynamic order. They will be described in Chapter 16. Here we shall cite an example of dynamic order, which manifests itself in laser radiation. The atoms of laser are excited by an external agency—by means of optical pumping. Each atom acts like an antenna, radiating a light pulse of the order of 3 m length, the radiation taking only 10^{-8} s. The emission of light by each atom occurs independently of the other atom. However, at a certain level of pumping there occurs a sharp transition to coherent radiation of very long light pulses (up to 300,000 km!) with a sharply enhanced intensity. The laser is itself an open system which is far from equilibrium and in which there arises a dynamic order—the self-organization of coherent emitters.

As we shall see further, dynamic order, the appearance of dynamic structures and their ordered behaviour in time are possible only when the system is far from equilibrium. Linear nonequilibrium thermodynamics, briefly described in this chapter, is valid only in some neighbourhood of equilibrium. Its basic propositions are expressed by relations (9.51) and (9.80). The first relation describes the coupling of various kinetic processes due to the difference of the off-diagonal coefficients L_{ij} $(i \neq j)$ from zero and the second is a mathematical expression of the Prigogine theorem of minimum entropy production in the stationary state. There is no doubt that coupled processes do occur in an open biological system. Therefore, the gen-

eral phenomenological theory worked out by Onsager and Prigogine allows one to account for important biological phenomena. The question of the applicability of the Prigogine theorem of minimum entropy production to biological systems is more complicated. As we have seen, the entropy production σ is minimum only in those stationary states of biological systems which are close to equilibrium. Such systems are described by linear relations (9.51). But in physics the linear dependence of the system's response on the factor responsible for this response is always only the first approximation, which holds for small influences. In our case, the "smallness" implies a small distance from equilibrium. In order to deal with biological systems and with their dynamic order, it is necessary to go beyond the range of linear thermodynamics.

Let us consider far-from-equilibrium systems. Three situations are encountered. First, the assumption of local equilibrium may be invalid, i.e., the Onsager relations $L_{ij} = L_{ji}$ are not fulfilled. Second, local equilibrium may be preserved, but the properties of the system are continuously changed with increasing distance from equilibrium. In such a case, the system retains some of the properties of linear systems, in particular the theorem of minimum entropy production in the stationary state remains to be valid. And, finally, in the third case, there appears a dynamic order: new types of organization of the substance in space and time which are displayed only by open far-from-equilibrium systems, which are called *dissipative structures.*

It is obvious that the appearance of dynamic order depends on the instabilities of the equilibrium and stationary states of the system. Let us consider the corresponding stability criteria.

If there is no entropy production in the system, then $d_i S = 0$ and in a special case of the evolution of heat dQ

$$dS = d_e S - dQ/T$$

where $d_e S$ is the entropy flow. If $d_i S \geqslant 0$, then

$$d_i S = dS - dQ/T \geqslant 0$$

Substituting into this expression the law of conservation of energy in the form

$$dQ = dE + p\,dV$$

where E is the energy of the system, we obtain the expression of the second law of thermodynamics:

$$T d_i S = T\,dS - dE - p\,dV \geqslant 0 \tag{9.116}$$

The stability condition for the equilibrium state has the form (at constant p and T):

$$\delta E + p\delta V - T\delta S \geqslant 0 \tag{9.117}$$

At constant S and V the variations $\delta E \geqslant 0$, i.e., the energy is at a minimum for stable equilibrium. At equilibrium $(\delta E)_{eq} = 0$ and $(\delta^2 E)_{eq} > 0$. On the contrary, for systems with constant E and V we have

$$\delta S \leqslant 0, \quad (\delta S)_{eq} = 0, \quad (\delta^2 S)_{eq} < 0 \tag{9.118}$$

that is, the entropy is maximum in the equilibrium state. In application to chemical reactions this means that

$$\left(\frac{\partial \mathcal{A}}{\partial \xi}\right)_{eq} (\delta\xi)^2 < 0$$

since

$$\Delta S = \int \sigma \, dt = \int \frac{\mathcal{A}}{T} v \, dt = \int_{\xi_{eq}}^{\xi} \frac{\mathcal{A}}{T} \, d\xi$$

Hence, the condition

$$\left(\frac{\partial \mathcal{A}}{\partial \xi}\right)_{eq} < 0$$

follows from Eq. (9.118). The general stability criterion for a set of chemical reactions may be represented in the form

$$\sum_{i,\,j} \left(\frac{\partial \mathcal{A}_i}{\partial \xi_j}\right)_{eq} \delta\xi_i \, \delta\xi_j < 0 \tag{9.119}$$

This thermodynamic condition is equivalent to the kinetic condition. At $t \to \infty$ the variation $\delta\xi \to 0$. Near equilibrium the fluctuations $\delta\xi$ satisfy the linear equations

$$\frac{d(\delta\xi_i)}{dt} = \sum_j a_{ij} \delta\xi_j \tag{9.120}$$

with the solutions

$$\delta\xi_i = (\delta\xi_i)_{t=0} \exp(\lambda t)$$

Here all the eigenvalues λ, which are the roots of the equation

$$\| a_{ij} - \lambda \delta_{ij} \| = 0$$

are real and negative, $\lambda < 0$, i.e., the fluctuations are monotonically damped (cf. page 335).

Near equilibrium, the affinity is expressed by the formula

$$\mathcal{A}_i \approx \sum_j \left(\frac{\partial \mathcal{A}_i}{\partial \xi_j}\right)_{eq} \delta\xi_j$$

since $(\mathcal{A})_{eq} = 0$. The dissipation function is equal to

$$\sigma = \sum_i v_i \frac{\mathcal{A}_i}{T} \equiv \sum_i \frac{d\xi_i}{dt} \frac{\mathcal{A}_i}{T} = \sum_i \frac{d(\delta\xi_i)}{dt} \frac{\mathcal{A}_i}{T} \geqslant 0$$

Substituting into this expression the values of $\mathcal{A}_i$ and $\delta\xi_i$, we obtain for each value of λ:

$$\lambda \sum_{i,j} \left(\frac{\partial \mathcal{A}_i}{\partial \xi_j}\right)_{eq} \delta\xi_i \, \delta\xi_j > 0$$

which coincides with Eq. (9.119), since $\lambda < 0$.

The stability condition for a nonequilibrium but stationary near-equilibrium state has already been considered above. For chemical reactions it has the form of Eq. (9.93):

$$\sum_i \delta v_i \, \delta \mathcal{A}_i \geqslant 0$$

Consider the chemical reaction

$$X + Y \longrightarrow C + D$$

The reactants and their concentrations are designated by the same symbols. We have

$$\mathcal{A} \propto \ln \frac{XY}{CD}, \quad v \propto XY$$

The fluctuation in the concentration X near the stationary state leads to an excess entropy production, which is proportional to

$$\delta v \, \delta \mathcal{A} \propto \frac{Y}{X} (\delta X)^2 > 0$$

The stability condition is satisfied.

We shall now deal with an *autocatalytic reaction* which proceeds with an increase in concentration of the reactant:

$$X + Y \longrightarrow 2X$$

As before, $v \propto XY$, but

$$\mathcal{A} \propto \ln \frac{XY}{X^2} = \ln \frac{Y}{X}$$

and

$$\delta v \, \delta \mathcal{A} \propto -\frac{Y}{X} (\delta X)^2 < 0$$

The stability condition is not fulfilled. But it remains valid near equilibrium. In order to find the equilibrium condition, one has to take into account the reverse reaction

$$X + Y \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} 2X$$

Then we have

$$v = k_1 XY - k_{-1} X^2, \quad \delta v = k_1 Y \delta X - 2k_{-1} X \delta X$$

At equilibrium $k_1 Y = k_{-1} X$ and we obtain

$$\delta v = -k_{-1} X \delta X \quad \text{and} \quad \delta v \delta \mathcal{A} \propto (\delta X)^2 > 0$$

Thus, far from equilibrium there can really arise unstable states of the dissipative system. The appearance of an instability in a certain initial state implies the transition of the system to a new regime, under which the system behaves in a different manner. Let us assume that there is a nonlinear system of chemical reactions, in the course of which the reactants A are converted to the final products F. The

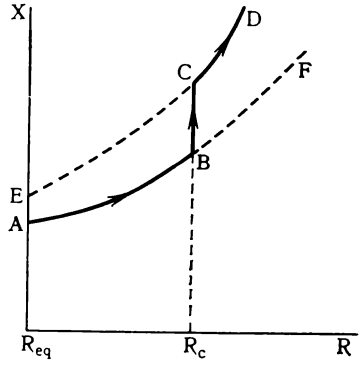


Fig. 9.4. The dependence of the steady-state concentration of the intermediate compound on the parameter R.

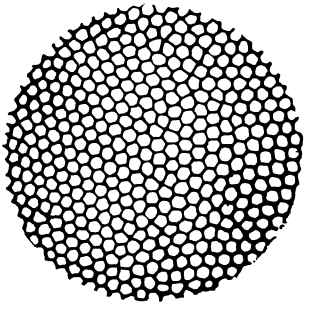


Fig. 9.5. The Bénard structures.

system may be characterized by a certain parameter R, which depends on the overall affinity, i.e., on the ratio of the concentrations A and F, and on the equilibrium constant. Figure 9.4 presents the steady-state concentration of the intermediate compound as a function of R. At small deviations from equilibrium $| R - R_{eq} |$ the system moves smoothly along the *thermodynamic* (*static*) *branch AB* in Fig. 9.4. All the stationary states on this branch are stable and conform to the theorem of minimum entropy production. However, at a sufficiently large distance from equilibrium, at a certain threshold value R_c, the excess entropy production equal to

$$\delta_X \sigma = \frac{1}{T} \sum_i \delta v_i \, \delta \mathcal{A}_i$$

may become negative. At this value of R_c there appears an instability, and the system moves to a new *kinetic* (*dynamic*) *branch CD*, on which the states are again stable. The regions of unstable states in Fig. 9.4 are shown by the lines *EC* and *BF*. On the kinetic branch

there may arise an organization of the system in time (say, undamped oscillations), an organization in space, and new multiple stationary states.

When the threshold R_c is attained, a transition of the type of a phase transition occurs as a result of the fluctuations being amplified up to a macroscopic level. A new regime and a new structure become stable. This effect is of course compatible with the boundary conditions imposed upon the system. Such is exactly the behaviour of the laser. A further example is the *Bénard instability*. A horizontal layer of a liquid heated from below passes, at a critical temperature gradient, to a state of internal convection, as a result of which ordered cooperative structures are formed (Fig. 9.5). This effect is directly associated with the nonlinearity of hydrodynamic equations. In a similar way, nonlinear autocatalytic chemical processes lead, in a region located beyond the thermodynamic threshold, to the formation of space-time dissipative structures (Chapter 16). The order in an equilibrium system appears in accordance with the condition of minimum free energy, and in a dissipative system it arises due to the increase of fluctuations up to a macroscopic level. The field of physics that deals with dissipative systems and with ordering is called *synergetics*. In this field we go beyond the area of thermodynamics and have to make recourse to concrete kinetic models (see Chapters 15-17). As we shall see, these models also prove to be effective in the study of biological development—*phylogenesis* and *ontogenesis* (Chapter 17).

Reference to dissipative systems is directly associated with the extension of information theory. It becomes necessary to study not only the amount of information, its transmission and recording but also the reception of information, which is possible only outside equilibrium, in the presence of instability. It is exactly on this path that the concept of the "anti-entropy nature" of life becomes physically meaningful (Chapter 17).

Katchalsky and his coworkers have developed a new method of deducing kinetic equations and of their analysis—the *network thermodynamics*. It is a phenomenological theory based on the analogy of the relationship between an electric circuit and an arbitrary dynamical system, be it mechanical, thermodynamical or chemical. The algorithm makes use of the theory of graphs. The thermodynamics of networks allows one to study nonlinear systems with feedforward and feedback connections consisting of discrete elements.

Chapter 10 Physics of Membranes

10.1. Cell Membranes

Noncellular forms of life do not exist on the earth. Viruses and bacteriophages cannot be regarded as independent self-sufficient living systems—of all the functions of the living cell they possess only the ability to transmit the genetic program. Conversely, the basic characteristics of life are inherent both in unicellular organisms and in the overwhelming majority of types of specialized cells of multicellular organisms. The structure and behaviour of individual cells are so complex that it proves possible to formulate the problems of behaviour at the cell level, the problems of *cytoethology* (Aleksandrov, 1970).

The building of ultramolecular (supermolecular) biology starts with the study of the living cell. The biophysics of the cell attempts to treat the cell through the use of experimental and theoretical models that are amenable to physical approaches.

The most important physical and physico-chemical functions of the cell consist of chemical metabolism and biosynthesis, in bioenergetic processes of storage of energy and its transformation during electro- and mechanochemical processes and regulated transport of molecules and ions. As we have seen (see Sec. 2.6), the storage of energy predominantly occurs in the form of ATP—the chemical energy of ATP is transformed into the chemical, electrical, osmotic, and mechanical work. The biosynthetic and bioenergetic functions are inextricably interlinked; they are realized only in an open nonequilibrium system. Accordingly, these functions are coupled to the transport of matter from the surrounding medium into the cell and vice versa.

The combination of the transport of matter with the preservation and independence of the inner cell structure is accomplished by the only possible way—for its functions to be realized the cell as a whole is isolated from its environment by a semipermeable partition. Each cell is surrounded by a plasma membrane. The appearance of a cell membrane was evidently an important stage in the origin of life—the *compartmentation*, the isolation of the intracellular space

from the outside world, determined the decisive acceleration of the prebiological and biological evolution (see Chapter 17).

On the other hand, delicate regulation of intracellular processes is accomplished on the basis of the spatial separation of cell organelles. The intracellular membranes serve as compartments for the internal content of the cell.

Biological membranes are supermolecular systems whose extent in two dimensions considerably exceeds their thickness, which is of the order of 10 nm. All the mechanisms responsible for the biological functionality of the membrane are, however, localized just in its thickness.

The membranes are not passive, semipermeable shells; they take a direct and very important part in all the functions of the cell. Biological membranes provide an *active transport* of matter moving against the opposing chemical or electrochemical potential gradient. The membranes serve as the site of the main bioenergetic processes—*oxidative phosphorylation* and *photosynthesis.* ATP is synthesized in *mitochondrial membranes* and in the *thylakoid membranes of the chloroplasts* of green plants. There is every ground for believing that there is a connection between the ribosomes, on which protein synthesis takes place, and the membrane system of the *endoplasmic reticulum* (EP). The replication of DNA and chromosomes presumably involves membranes.

The most important bioenergetic processes include bioelectric phenomena—the generation of *biopotentials.* The propagation of the nervous impulse is a membrane process. Reception, whether by touch, smell, taste, hearing, or vision, occurs with the invariable participation of membranes.

From what has just been said it follows that the physics of membranes is one of the central areas of biophysics. Membrane physics and biology are of fundamental significance to theoretical science and to its applications in medicine and pharmacology.

Below is given a list of the basic problems of the physics of membranes.

1. The membrane structure and the dynamical properties of the membrane structure, which determine its functionality (Sections 10.2 and 10.3).
2. The passive and active membrane transport (Sections 10.4, 10.5, and 10.6).
3. The properties of excitable membranes (Chapters 11 and 12).
4. The membrane bioenergetics (Chapters 12, 13, and 14).
5. The physics of reception processes (Sections 11.7, 12.6, and 14.6).

In concluding this section, we give a table which lists the most important events occurring in biological membranes of various types. The quantum $\hbar\omega$ indicates the phenomena associated with light quanta; the quantity $\Delta\psi$ represents the change of the electrical po-

tential; i denotes the appearance of an electric (ionic) current; e represents the transport of electrons; ATP (+) is the synthesis of ATP; ATP (—) is the degradation of ATP.

Table 10.1. The Events Taking Place in Membranes (after Witt)

Photosynthesis	$\hbar\omega$	$\Delta\psi$	i	e	ATP (+) and (—)
Respiration	—	$\Delta\psi$	i	e	ATP (+) and (—)
Vision	$\hbar\omega$	$\Delta\psi$	i	—	ATP (—)
Nerves	—	$\Delta\psi$	i	—	ATP (—)
Muscles	—	$\Delta\psi$	i	—	ATP (—)

10.2. The Structure of Membranes

A biological membrane is a dynamical organized system; it is necessary to study both its structure and the dynamics of its behaviour.

Biological membranes consist mainly of lipids and proteins. The cells of the mammals also contain small amounts of carbohydrates associated with proteins (glycoproteins) or with lipids. The inner membranes largely contain phospholipids, and the plasma membranes contain also neutral lipids. For example, in the membranes of erythrocytes 30 percent of lipids is present in the form of *cholesterol.*

The extraction of individual components from membranes is effected with the aid of *detergents* (say, sodium dodecylsulphate), which solubilize insoluble substances, and through the separation of the proteins produced by means of electrophoresis in the polyacrylamide gel.

In most cases, membranes are heterogeneous. They contain whole families of phospholipids and lipids. For example, the membranes of human erythrocytes contain no less than 20 types of lecithin. Lipids are built up of a polar "head" and two long nonpolar hydrocarbon "tails", which possess hydrophobic properties.

The proteins of membranes are also varied. About one-third of the proteins of the erythrocyte membrane is spectrin composed of two components with molecular weights 255,000 and 220,000. The second are a series of proteins with molecular weights of about 90,000, and the third are proteins with a molecular weight of 9000-15,000. Membranes of simpler composition also exist—the inner membranes of the rods of retina contain only one protein, *rhodopsin* (see Sec. 14.7).

As far back as 1935 Danielli and Davson proposed a theoretical model of the composition of biological membranes—the *unit membrane.* The unit membrane is considered to be a bilayer of mixed polar lipids, with their hydrophobic tails extending inward and their hydrophilic heads exposed on the surface where they are bound to the outer monomolecular protein layers (Fig. 10.1).

The main source of information about the cell structure and cell membranes is electron microscopy. To obtain electron micrographs, the specimens are stained with OsO_4, $KMnO_4$, etc. The chemistry of the processes involved has not been studied in detail; it is also unclear what happens upon extraction of membranes and preparation of specimens. Artifacts are not excluded. Nonetheless, the main idea of the organization of the unit membrane—the bilayer arrangement of lipids—is correct. This is also proved by X-ray data. As regards their protein coat, the arrangement of proteins differs from that suggested in the symmetrical unit membrane model, according to which the membrane proteins have hydrophilic surfaces, which become bound to the polar end of the lipids. Isolated membrane proteins and membranes have also been studied by indirect methods—by the action of proteolytic enzymes on them and by incorporation of various labels into the membrane proteins. It has been found that the membrane proteins can be subdivided into two classes. Some of the proteins are bound only to the membrane surfaces; like globular proteins, which function in an aqueous environment, they have a hydrophilic surface. The proteins of the second class are capable of penetrating the membrane and interacting with the hydrophobic tails of the lipids. These proteins are insoluble in water, their surfaces being hydrophobic. The investigations of the membranes carried out by means of infrared spectroscopy, spectropolarimetry, NMR, etc. point to the diversity of protein structures and to interprotein interactions, which are not taken into account in the unit membrane model. It has been established that the proteins are arranged asymmetrically in the membranes (Bergelson, 1970).

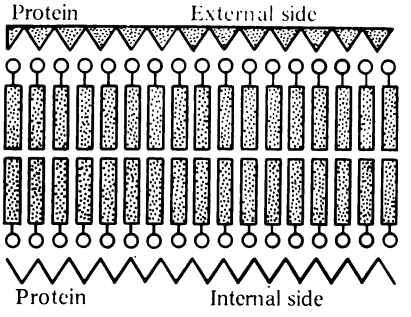


Fig. 10.1. The model of the biological membrane. The circles are the polar heads of lipids and the rectangles are the nonpolar fatty-acid tails.

Important results have been obtained by the *freeze-etching* method in which the membranes are rapidly frozen at the liquid-nitrogen temperature and "fractured" with a cold, chisel-like tool in a vacuum. The ice sublimates and the specimen is "shadowed", replicated with platinum and carbon and examined under an electron microscope.

It has been found that the cleaved surface passes along the internal hydrophobic portion of the red cell membrane. Globular particles

with a diameter of up to 7.5 nm are found to exist in the membranes. These particles are proteins.

There are several modifications of the unit membrane model. At present, the model that is thought to be closer to the truth is the *fluid mosaic model* of the cell membrane shown in Fig. 10.2. This model too has a bilayer of lipids. As a matter of fact, *artificial lipid membranes* with a bilayer structure are found to be similar to biological membranes in many respects. Artificial membranes are

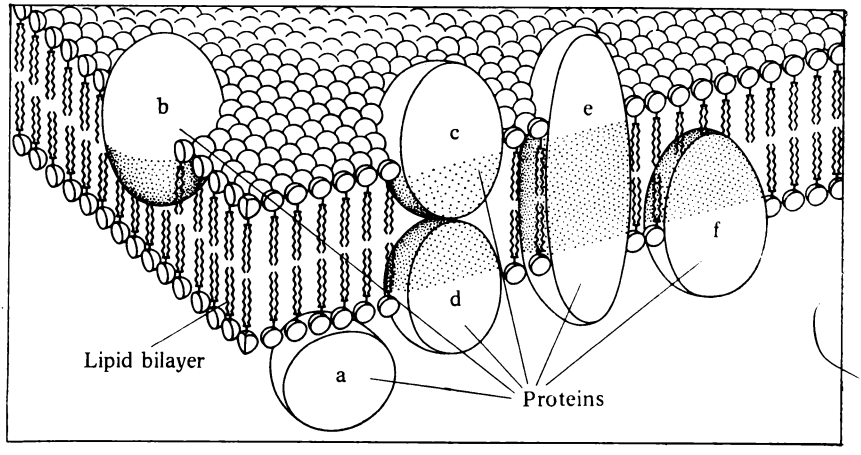


Fig. 10.2. A present-day schematic representation of the structure of the cell membrane: (*a*) through (*f*) are different protein molecules.

produced when a mixture of phospholipids and neutral lipids dissolved in organic solvents is allowed to contact with water. In this way, "black" membranes can be produced, i.e., thin layers 10 nm thick which are devoid of interference colours. It has been found that these membranes have the bilayer structure indicated. Table 10.2 compares the properties of artificial bilayer lipid membranes with those of biological membranes.

Artificial membranes display no metabolic activity and are not so highly selective as are biological membranes. But they simulate the important properties of biomembranes and enable one to study the transport of material and excitability.

It has been established that various complex liquid-crystalline structures of long-range order are formed in lipid-water systems. These systems are of scientific interest in their own right. There is

Table 10.2. Comparison of the Properties of Bilayer Lipid Membranes with Those of Biological Membranes

Property	Biological membranes at 25°C	Bilayer membranes at 36°C
Thickness, nm	6-10	6.7-7.5
Electric capacitance, pF/mm^2	0.5-1.3	0.38-1.0
Electric resistance, ohm·cm^2	10^2-10^5	10^6-10^9
Breakdown voltage, mV	100	150-200
Surface tension, N/cm^2	$0.03\text{-}1.0 \times 10^{-5}$	$0.5\text{-}2.0 \times 10^{-5}$
Water permeability, μm/s	0.37-400	31.7
Activation energy for transport of water, kJ/mole	40.3	53.3

ground for believing that the multiphase nature of lipid systems is directly associated with the functions of biological membranes (see further in the text).

10.3. Conformational Properties of Membranes

Let us consider the dynamical properties of membranes. A number of facts point to the high mobility of the lipid bilayer. The lipids in the membrane behave like *liquid crystals*. It is precisely a liquid crystal that combines high ordering with fluidity and lability. This combination makes possible for membranes to perform their important functions.

The liquid-crystalline (liquid) properties of biological membranes are determined by the fact that the lipids present in them are in a molten state at the physiological temperature. The melting temperature of a hydrocarbon is the lower, the larger is the number of double bonds it contains (this gives rise to a difference between animal fats and vegetable oils). The lipids that have double bonds in their hydrocarbon chains melt at temperatures lower than the physiological temperature. The fraction of such lipids in the plasma membranes of mammals is large.

The liquid properties of membranes are proved by many facts. The mobility of the membrane structure is revealed with the aid of paramagnetic and fluorescent labels and also by means of the NMR method. Figure 10.3 shows the EPR spectrum of a membrane labeled with nitroxyl attached to the lipid tails. At a low temperature the lipid is frozen and at a high temperature the spectrum is peaked since the lipid has melted.

Especially thoroughly studied are the liquid-crystalline properties of photoreceptor membranes containing the protein rhodopsin (Sec. 14.7). There is one molecule of rhodopsin in the membrane per 60-90 lipid molecules, of which 80 percent contain an unsa-

turated fatty acid. The flash photometric technique has revealed that the rhodopsin molecule rotates rapidly about the axis perpendicular to the membrane plane. The time of such rotational diffusion is 20 μs at 20°C. The study of the bleaching of rhodopsin in the light by the microspectrophotometric technique has shown that the rhodopsin in the membrane undergoes translational lateral diffusion. The diffusion coefficient is equal to $(3.5 \pm 1.5) \times 10^{-9}$ $cm^2 \cdot s^{-1}$, which corresponds to the viscosity from 0.1 to 0.4 P. A close value is shown by the viscosity of the cell membranes of mammals determined from translational diffusion and also by the viscosity of the membranes of mitochondria and nerve axons. Thus, the membrane viscosity is two or three orders of magnitude higher than the viscosity of water and corresponds to the viscosity of a vegetable oil. More viscous membranes are also known to exist.

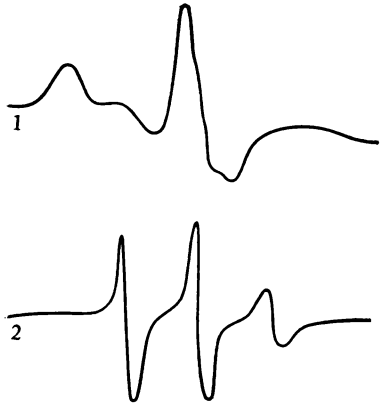


Fig. 10.3. The EPR spectra of the spin-labeled lipid molecule in the membrane:
1—low temperature; *2*—high temperature.

The membrane structure in Fig. 10.2 shows masses of globular proteins "floating" in an oily layer of lipid. Their molecules are embedded on both sides of the membrane to varying degrees in the double bilayer of mobile hydrocarbon tails of the lipids. There are proteins that span the entire membrane thickness. A considerable part of the membrane surface is free from proteins: for example, proteins occupy 70 percent of the membrane surface in red blood cells and 80 percent of the surface of the microsome membrane.

There are *channels* in membranes through which small ions and molecules are transported. Presumably, these channels are formed in the neighbourhood of protein molecules; they are coated with polar groups. The nature of the channels constitutes a topical problem of the physics of membranes (see Sec. 11.4).

A number of facts indicate conformational transitions in membranes. Structural changes are detected with the aid of fluorescent and paramagnetic labels, by measuring birefringence and light scattering, and by circular dichroism. In membranes there are observed phase transitions—the melting of lipids. Such a transition occurs near 0°C when the membranes of mitochondria and microsomes are heated from —40°C. With the aid of spin-labels in a suspension of

plasma membranes isolated from the mouse fibroplast there have been established the temperatures of the lateral phase separation in lipids. For the outer monomolecular lipid layer such transitions are observed at 15 and 31°C. and for the inner layer they are detected at 21 and 37°C.

The mobility of membrane lipids and phase transitions that occur in them are determined by their conformational properties. The melting of lipids occurs through the rotational isomerization of

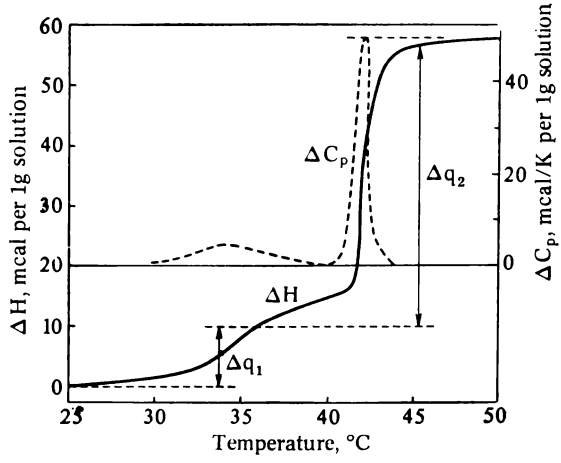


Fig. 10.4. The temperature dependence of the enthalpy and heat capacity of a solution of dipalmitoyl-α-lecithin (Δq_1 and Δq_2 are the heats of phase transitions).

hydrocarbon chains—this is a conformational melting. The saturated hydrocarbons (paraffins) crystallize in the form of a continuous series of *trans*-rotamers (see page 70). In addition to *trans*-rotamers, the melting gives rise to *gauche*-rotamers. Their fraction in liquid paraffinic hydrocarbons is about 10 percent. This also refers to the hydrocarbon tails of the lipids.

Figure 10.4 shows the manner in which the heat capacity and enthalpy of a solution of the lipid dipalmitoyl-α-lecithin change with temperature. Two phase transitions are observed, one at 34°C and the other, especially sharp one, at 41°C. The X-ray pattern taken at a temperature lower than the transition temperature shows sharp diffraction rings corresponding to interchain distances of 0.48 nm. At a temperature higher than the transition temperature there is observed a diffraction ring corresponding to an interchain distance of 0.53 nm.

There has been worked out the statistical mechanics of bilayer phospholipid membranes which takes account of rotamerization and interchain steric restrictions. If the energy difference between the rotamers is assumed to be equal to $\Delta E \approx 2.5$ kJ/mole (cf. page 70), the enthalpy and entropy changes brought about by the melting of a bimolecular lipid layer are found to be in agreement with experiment.

Träuble (1971) advanced the theory of transport of molecules across the lipid membrane based on the same conceptions of the rotamerization of hydrocarbon chains. The staggered (*gauche-*) rotamers are considered to be mobile structural defects (kinks) that determine the presence of free volumes in the hydrocarbon phase of the membrane. The diffusion coefficient of such kinks is estimated as 10^{-5} cm$^2 \cdot$s^{-1}; this is fast diffusion. The transport of small molecules occurs as a result of their entry into free volumes and their joint migration. This idea provides a model of "kinetic channels" in the membrane and permits one to calculate its permeability to water in agreement with experiment.

Conformational changes play an important role in the interactions of membranes with various ligands, which is essential to physiology and pharmacology.

In most cases, the response of cell membranes to the addition of specific ligands is cooperative in character. The curves of the response of the membrane and cell to the increasing concentration of ligand often display inflections (cf. page 218).

The so-called colicinogenic strains of the bacteria *Escherichia coli* produce macromolecular antibiotics—*colicins* which are capable of destroying the bacteria of other, "sensitive" strains of *E. coli*. In this process, the number of colicin molecules required for the destruction of a single bacterium may be very small; it may even be equal to unity. Evidently, the membrane of a sensitive cell has amplifying properties—the reception of one molecule serves as a *trigger*, bringing about macroscopic events on the cell scale.

In principle, similar trigger processes are presumably realized in the membranes of receptor cells (see Sec. 11.7).

It has been found that many drugs have an effect on the conformations of membranes and membrane lipids. Changeux and his coworkers treated the cell membrane as an ordered cooperative system built up of interacting subunits. They considered the triggering properties of membranes on the basis of a theory analogous to the theory of indirect cooperativity of enzymes developed by Monod, Wyman, and Changeux (see Sec. 6.7). Each subunit has a receptor site for a given specific ligand, the affinity for which is changed when its conformation is altered. In the ordered "lattice" of the membrane the subunits (protomers) interact with their neighbours, which is what gives rise to cooperative properties. Depending on the ligand

activity and the interaction energy of protomers, the response of the membrane to ligand addition may be gradual or sigmoidal, becoming in the limit an "all-or-none" transition—a phase transition. The formal model describes the action of colicins and provides a qualitative explanation for a number of facts, in particular, for the fact that various related drugs cause different maximum responses of the membranes. The primary action of many drugs is localized in membranes and is cooperative. Many drugs are effective in very small concentrations (down to 10^{-11} M) and are highly specific. The effect of a drug on the membrane receptor is determined by molecular recognition, but we still know little about the nature of these receptors (see Sec. 11.7).

The main difficulty in building a molecular theory of membrane transport and reception consists in analysing the dynamic interaction between proteins and lipids. Membrane receptors—presumably, proteins (rhodopsin in photoreceptors)—change their conformation when bound to the ligand, which leads to a change in depth of immersion and mobility of proteins in the "lipid sea". The factor responsible for cooperativity may be the interaction of "floating" proteins upon their collisions. The fluid mosaic model may serve as the basis for the molecular physics of membranes.

We may suppose that the properties of membranes are largely determined by electronic-conformational interactions (ECI, see Sections 6.6 and 13.4). The local shift of electron density that arises during the interaction of the membrane receptor with the ligand leads to conformational changes in biological molecules. The transport of electrons and ions across the membrane may be treated as the propagation of *conformons*—quasi-particles consisting of carriers of electronic charge, or as an electron-density shift and conformational displacements of the surrounding medium (see pages 215, 457, and 460).

10.4. Passive Membrane Transport

The consideration of membrane transport should naturally be started with thermodynamics, without reference to the molecular structure of the membrane and transport mechanism. Thermodynamics not so much explains physical phenomena as organizes our knowledge, establishing the relationship between physical phenomena and their dependence on the parameters of the system.

In this particular case, thermodynamics reveals the dependence of transport fluxes on the averaged characteristics of solutions and membranes.

Consider the simplest model—a homogeneous membrane which is in contact with two solutions (Fig. 10.5). The dissipation function inside the element dx of the membrane may be written in the

form

$$T\sigma = \sum J_i \nabla (-\mu_i) \tag{10.1}$$

where J_i is the vector flux of the ith component through the membrane; μ_i is the chemical potential. Integrating this expression with respect to the membrane thickness, we obtain

$$\Psi = \int_0^{\Delta x} \sigma T \, dx = \sum_i J_i [\mu_i (0) - \mu_i (\Delta x)] = \sum_i J_i \Delta \mu_i \tag{10.2}$$

If there is one solute (s) and one solvent (water, denoted as w), then

$$\Psi = J_s \Delta \mu_s + J_w \Delta \mu_w \tag{10.3}$$

Calculations show that this integrated dissipation function may be rewritten as follows:

$$\Psi = J_v \Delta p + J_D \Delta \pi \tag{10.4}$$

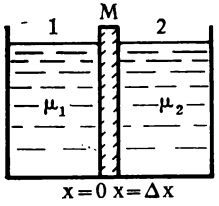


Fig. 10.5. Illustrating the membrane transport.

where Δp is the difference in pressure on both sides of the membrane; $\Delta \pi$ is the difference in osmotic pressure; J_v is the total volume flow:

$$J_v = J_s V_s + J_w V_w \tag{10.5}$$

The quantities V_s and V_w are the partial molar volumes; J_D is the flow of the solute relative to water:

$$J_D \approx v_s - v_w \tag{10.6}$$

The quantities v_s and v_w are the rates at which the solute and water are transported across the membrane. With a large excess of solvent

$$J_v \approx v_w \tag{10.7}$$

Thus, we have now passed over to generalized fluxes J_v and J_D and generalized forces Δp and $\Delta \pi$. Near equilibrium we have (see Sec. 9.3):

$$\begin{aligned} J_v &= L_p \Delta p + L_{pD} \Delta \pi \\ J_D &= L_{pD} \Delta p + L_D \Delta \pi \end{aligned} \tag{10.8}$$

The fluxes J_v and J_D are coupled. The coefficient L_p characterizes the *mechanical filtration capacity* of the membrane, i.e., the velocity of the liquid per unit pressure difference; L_D represents the velocity of the liquid per unit osmotic pressure difference; $L_{pD} = (J_v)_{\Delta p = 0} / \Delta \pi$ is the *osmotic flow coefficient*. For an ideal semipermeable membrane,

which is impermeable to the solute, $J_s = 0$ and $J_v = -J_D$. From this it follows that

$$(L_p + L_{pD})\,\Delta p + (L_{pD} + L_D)\,\Delta\pi = 0$$

and

$$-L_{pD} = L_p = L_D, \quad \Delta\pi = \Delta p$$

For a nonideal membrane the measure of its *selectivity* is

$$\varkappa = -\frac{L_{pD}}{L_p} = -\left(\frac{J_D}{J_v}\right)_{\Delta\pi=0} = \left(\frac{v_w - v_s}{v_w}\right) < 1 \tag{10.9}$$

or

$$\varkappa = \left(\frac{\Delta p}{\Delta\pi}\right)_{J_v=0} \tag{10.10}$$

The quantity $\varkappa$ is called the *reflection coefficient*. For the membrane of the human erythrocyte $\varkappa = 0.62$ with respect to urea.

The phenomenological coefficients L_p, L_D, and L_{pD} may be expressed through the friction coefficients of water and solute upon contact with the membrane and of the solute upon contact with water.

If the solute consists of charged ions (of importance to biology are mainly the ions K^+, Na^+, Ca^{2+}, and Mg^{2+}), then, instead of the chemical potential, one must consider the *electrochemical potential* defined as

$$\tilde{\mu}_j = \mu_i + z_i \mathcal{F} \psi \tag{10.11}$$

where ψ is the electrical potential of the membrane; z_i is the valency of the charged particle; $\mathcal{F}$ is Faraday's number. When the two solutions (1 and 2) separated by the membrane are at equilibrium, their electrochemical potentials are equal:

$$\tilde{\mu}_i^{(1)} = \tilde{\mu}_i^{(2)} \tag{10.12}$$

or

$$\Delta\mu_i = RT \ln \frac{a_i^{(1)}}{a_i^{(2)}} = -z_i \mathcal{F} \Delta\psi \tag{10.13}$$

where $a_i^{(1)}$ and $a_i^{(2)}$ are the activities, which in the case of ideal solutions are replaced by the concentrations $c_i^{(1)}$ and $c_i^{(2)}$.

Phenomenological nonequilibrium thermodynamics as applied to *passive ion transport* is built up by analogy with the description of the transport of neutral molecules. The phenomenological coefficients are also expressed through the friction coefficients. The situation here is more complicated, since the number of these coefficients is large—there are six coefficients for a solution of NaCl in water. Calculations are simplified if the membrane is strongly charged, and therefore the concentration of fixed counter-ions in the membrane is much higher than the concentration of the neutral salt.

Nonequilibrium thermodynamics offers a physically substantiated description of passive transport. It determines the kinetic characteristics of the membrane (say, $\varkappa$), which can be measured experimentally. We see that the treatment of membrane permeability requires a study of nonequilibrium fluxes of material. The transport dynamics is tied up with the properties of the membrane. The experimentally found linear dependence of fluxes on generalized forces (the gradients μ and $\tilde{\mu}$) for a number of passive artificial and

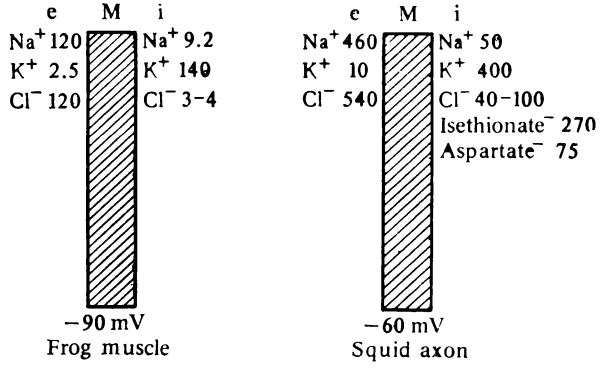


Fig. 10.6. The concentration of ions (mmole/litre) and the potential difference between the two sides of the cell membrane.

biological membranes implies the nearness of the coupled processes under consideration to equilibrium. Let us now turn to the molecular picture.

Living cells are characterized by the fact that the concentration of K^+ ions inside the cells may be 10-20 times higher than in the surrounding medium. For Na^+ ions there are observed gradients of the same order but opposite sign. Experiments with labeled atoms show that the ions of the cytoplasm are exchanged for the ions of the environment, i.e., the cell membrane is permeable to K^+ and Na^+ ions. Associated with the presence of ion concentration gradients is the experimentally observed potential difference of 50-70 mV between the cytoplasm and the environment. Figure 10.6 shows these relationships for the frog muscle and the squid axon.

Why are Na^+ ions expelled from the cell and K^+ ions left in an excess concentration? This question is in the sphere of evolution. The first cells presumably originated in sea water (see Sec. 17.1) and the composition of the intercellular medium, say the blood

plasma, is close to the composition of sea water. In order to create an electrochemical potential on the cell membrane, which is required for a number of biological functions, at the expense of an excess of Na^+ ions, there would be required concentrations of Na^+ ions of the order of several moles per litre inside the cell. Conversely, the amount of K^+ ions in the medium (in sea water) is so small that the required potential is produced with intracellular concentrations less by an order of magnitude.

Experiments show that both K^+ and Na^+ ions move practically freely in the axoplasm. That is why their concentration gradients are determined by the special properties of the membrane rather than by the specificity of the cytoplasm. The resting potential, i.e., the potential difference for the unexcited membrane, characterizes a certain nonequilibrium steady-state process.

We shall first consider the ionic equilibrium between two solutions separated by the membrane. For example, for NaCl we have

$$\mu^i_{Na} + \mu^i_{Cl} = \mu^e_{Na} + \mu^e_{Cl}$$

(the subscripts i and e stand for the internal and external sides of the membrane) or

$$\Delta\mu_{Na} = \mu^e_{Na} - \mu^i_{Na} = -(\mu^e_{Cl} - \mu^i_{Cl}) = -\Delta\mu_{Cl}$$

Since $\Delta\mu$ depends logarithmically on the ratio of the activities [cf. Eq. (10.13)], it follows that

$$\frac{a^i_{Na}}{a^e_{Na}} = \frac{a^e_{Cl}}{a^i_{Cl}} \tag{10.14}$$

This is the *Donnan equilibrium*. The activity ratio may be approximately replaced by the concentration ratio.

If the permeability of the membrane to both species of ions is determined only by their mobility, then with a difference in concentrations there will arise a diffusion potential on both sides of the membrane which is determined by the high mobility of Cl^- ions as compared to that of Na^+ ions. Therefore, a more dilute solution will become electronegative with respect to a more concentrated solution. The potential difference will be given by [cf. Eq. (10.13)]

$$\Delta\psi \quad \frac{RT}{\mathcal{F}} \frac{u_{Na} - u_{Cl}}{u_{Na} + u_{Cl}} \ln \frac{c_1}{c_2} \tag{10.15}$$

where $u_{Na} = 5.2$ and $u_{Cl} = 7.9$ (μm/s) (cm/V) are the mobilities of the ions. If the ratio of the concentrations on both sides of the membrane is $c_1/c_2 = 10$, then $\Delta\psi = -12$ mV. If the membrane is permeable only to Na^+ ions, then $u_{Cl} = 0$ and we obtain the

Nernst equation, which coincides with Eq. (10.13):

$$\Delta\psi = \frac{RT}{\mathcal{F}} \ln \frac{c_1}{c_2} \tag{10.16}$$

At $c_1/c_2 = 10$ the potential difference is $\Delta\psi = 60$ mV.

In biology we deal with a more complex system—we must take into account the presence of at least three species of ions: K^+, Na^+, and Cl^-.

Assuming the electric field to be constant and uniform throughout the entire thickness of the membrane, Hodgkin and Katz (1949) derived a formula for the resting membrane potential $\Delta\psi$ which is treated as the diffusion potential. The currents of univalent ions flowing across the membrane are given by

$$\begin{aligned} I_K &= -RTu_K \frac{dc_K}{dx} - c_K u_K \mathcal{F} \frac{d\psi}{dx} \\ I_{Na} &= -RTu_{Na} \frac{dc_{Na}}{dx} - c_{Na} u_{Na} \mathcal{F} \frac{d\psi}{dx} \\ I_{Cl} &= RTu_{Cl} \frac{dc_{Cl}}{dx} - c_{Cl} u_{Cl} \mathcal{F} \frac{d\psi}{dx} \end{aligned} \tag{10.17}$$

For a uniform field and a uniform membrane

$$\frac{d\psi}{dx} = -\frac{\Delta\psi}{\Delta x} \tag{10.18}$$

where Δx is the thickness of the membrane. Substitution of Eq. (10.18) into the first equation in (10.17) and integration from c_K^i to c_K^e gives

$$I_K = \frac{u_K \mathcal{F} \Delta\psi}{\Delta x} \frac{c_K^e - c_K^i \exp(-\mathcal{F}\Delta\psi/RT)}{1 - \exp(-\mathcal{F}\Delta\psi/RT)} \tag{10.19}$$

The permeability coefficient of the membrane for K^+ ions is given by

$$P_K = \frac{u_K RT}{\Delta x \mathcal{F}} \tag{10.20}$$

From Eqs. (10.19) and (10.20) it follows that

$$I_K = \frac{P_K \mathcal{F}^2 \Delta\psi}{RT} \frac{c_K^e - c_K^i \exp(-\mathcal{F}\Delta\psi/RT)}{1 - \exp(-\mathcal{F}\Delta\psi/RT)} \tag{10.21}$$

The net current is equal to the sum of the currents I_K, I_{Na}, and I_{Cl}:

$$I = \frac{P_K \Delta\psi \mathcal{F}^2}{RT} \frac{w - y \exp(-\mathcal{F}\Delta\psi/RT)}{1 - \exp(-\mathcal{F}\Delta\psi/RT)} \tag{10.22}$$

where

$$w = c_K^e + \frac{P_{Na}}{P_K} c_{Na}^e + \frac{P_{Cl}}{P_K} c_{Cl}^i$$

$$y = c_K^i + \frac{P_{Na}}{P_K} c_{Na}^i + \frac{P_{Cl}}{P_K} c_{Cl}^e$$

If the ionic current I is very small, i.e., if the resistance of the membrane is high, then $I \approx 0$, $w \approx y \exp(-\mathcal{F}\Delta\psi/RT)$, whence

$$\Delta\psi = \frac{RT}{\mathcal{F}} \ln \frac{y}{w}$$

or

$$\Delta\psi = \frac{RT}{\mathcal{F}} \ln \frac{P_K c_K^i + P_{Na} c_{Na}^i + P_{Cl} c_{Cl}^e}{P_K c_K^e + P_{Na} c_{Na}^e + P_{Cl} c_{Cl}^i} \qquad (10.23)$$

We have deduced the *Hodgkin-Katz formula.* It agrees well with experiment at large c_K^e and small c_K^i. When c_K^e decreases or c_K^i increases, the value of the quantity $\Delta\psi$ tends to a certain limit. The product $P_K c_K^i$ behaves as a constant; an increase in the resting potential $\Delta\psi$ leads to a decrease in the permeability coefficient P_K. Theoretical models have been constructed to account for these facts which take account of the necessity of overcoming the activation barriers by the ions that traverse the membrane. We shall return to these questions in Sections 10.6 and 11.3.

10.5. Active Membrane Transport

An increased concentration of K^+ ions and a decreased concentration of Na^+ ions inside the cell are determined by the active membrane transport which proceeds against the electrochemical potential gradient.

Active transport is one of the most important features of life processes. It resolves the contradiction between the preservation of spatial heterogeneity and metabolism—the exchange of matter and energy with the surrounding medium.

Active transport is accomplished through the coupling of diffusion fluxes to the exergonic reactions that take place in the bulk of the membrane. The transfer of matter occurs at the expense of the free energy liberated in chemical reactions. As a rule, this is the energy of hydrolysis of ATP. The coupling indicated is not trivial. As has already been pointed out (see page 324), the coupling coefficients for scalar and vector fluxes are equal to zero in an isotropic system in accordance with the Curie principle. The coupling of chemical reactions to diffusion may be indirect, arising as a result of the stationary state being maintained (page 336). On the

other hand, direct coupling is also effected due to the anisotropy of the membrane. Finally, there is a transport facilitated by a chemical reaction (known as *facilitated* or *mediated transport*). In this case, the transport is accelerated due to the presence in the membrane of *carriers*—substances which interact with the transported ions or molecules. Ions or molecules form a complex with the carrier on the outer side of the membrane. This complex is broken down on the inner side. Neither the carrier (C) nor its complex with the transported substance (S) leaves the membrane. The following reaction takes place inside the membrane:

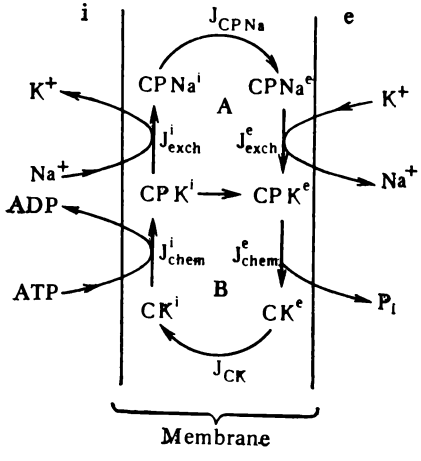


Fig. 10.7. The thermodynamic scheme of the sodium pump.
The superscripts *i* and *e* refer to the internal and external sides of the membrane; *A*—the ion-exchange cycle; *B*—the chemical cycle.

$$S+C \rightleftharpoons CS$$

A flux of reactant S passes through the membrane. From the stationary-state condition it follows that the flux of the carrier in the membrane J_C is compensated for by the opposing flux of the complex J_{CS}:

$$J_C+J_{CS}=0$$

that is, the circulation of C occurs.

Let us consider, in this connection, the *sodium pump*—a transporting mechanism operative in the membrane (in particular, in the membrane of the nerve cell) which makes use of the free energy of ATP for the active transport of Na^+ and K^+ ions in the direction of their increasing concentrations. The key role in the operation of the sodium pump is the enzyme—the K^+, Na^+-*activated ATPase*. We shall describe the sodium pump in terms of carriers, which can be phosphorylated and dephosphorylated in the various regions of the membrane:

$$C+ATP \rightarrow CP+ADP$$

Suppose that ATP and ADP react only on the inner side of the membrane. The phosphorylated carrier CP undergoes dephosphorylation as a result of a certain conformational conversion:

$$CP \xrightarrow{H_2O} C+P$$

Let us assume that the carrier C is a protein with high affinity for K^+, and CP is a protein which predominantly binds Na^+. The sodium pump operates as a system of two cycles, in which one cycle drives the other (Fig. 10.7). The first cycle is of the ion-exchange type:

$$CPK + Na^+ \rightleftharpoons CPNa + K^+$$

The second cycle is chemical and involves phosphorylation and dephosphorylation reactions.

In the stationary state

$$\frac{dCK^i}{dt} = -J^i_{chem} + J_{CK} = 0, \quad \frac{dCK^e}{dt} = J^e_{chem} - J_{CK} = 0$$

that is,

$$J^e_{chem} = J^i_{chem} = J_{CK} \tag{10.24}$$

For the chemical cycle we obtain

$$\frac{dCPNa^i}{dt} = J^i_{exch} - J_{CPNa} = 0, \quad \frac{dCPNa^e}{dt} = J_{CPNa} - J^e_{exch} = 0$$

that is,

$$J^i_{exch} = J^e_{exch} = J_{CPNa} \tag{10.25}$$

and

$$\begin{aligned} \frac{dCPK^i}{dt} &= J^i_{chem} - J^i_{exch} - J_{CPK} = 0 \\ \frac{dCPK^e}{dt} &= J_{CPK} + J^e_{exch} - J^e_{chem} = 0 \end{aligned} \tag{10.26}$$

that is,

$$J_{CPK} = J^i_{chem} - J^i_{exch} = J^e_{chem} - J^e_{exch} \tag{10.27}$$

The dissipation function has the form

$$T\sigma = J_{CK}\Delta\mu_{CK} + J^i_{chem}\mathcal{A}^i_{chem} + J^e_{chem}\mathcal{A}^e_{chem} + J_{CPK}\Delta\mu_{CPK} + \\ + J_{CPNa}\Delta\mu_{CPNa} + J^i_{exch}\mathcal{A}^i_{exch} + J^e_{exch}\mathcal{A}^e_{exch} \tag{10.28}$$

where

$$\Delta\mu_{CK} = \mu^e_{CK} - \mu^i_{CK}, \quad \Delta\mu_{CPK} = \mu^e_{CPK} - \mu^i_{CPK}, \quad \Delta\mu_{CPNa} = \mu^i_{CPNa} - \mu^e_{CPNa}$$

$$\mathcal{A}^i_{chem} = \mu_{ATP} + \mu^i_{CK} - \mu_{ADP} - \mu^i_{CPK}$$

$$\mathcal{A}^e_{chem} = \mu^e_{CPK} + \mu_{H_2O} - \mu^e_{CK} - \mu_P$$

$$\mathcal{A}^i_{exch} = \mu^i_{CPK} + \mu^i_{Na} - \mu^i_{CPNa} - \mu^i_K$$

$$\mathcal{A}^e_{exch} = \mu^e_{CPNa} + \mu^e_K - \mu^e_{CPK} - \mu^e_{Na}$$

Hence,

$$T\sigma = J_{chem}(\mu_{ATP} + \mu_{H_2O} - \mu_{ADP} - \mu_P) + \\ + J_{exch}[(\mu^i_{Na} - \mu^e_{Na}) - (\mu^i_K - \mu^e_K)] = J_{chem}\mathcal{A}_{chem} + J_{exch}\mathcal{A}_{exch} \tag{10.29}$$

The ion-exchange coefficient is equal to

$$\Gamma = \frac{[K^i][Na^e]}{[K^e][Na^i]} \tag{10.30}$$

The phenomenological relations are as follows:

$$\begin{aligned} J_{exch} &= L_{11}\mathcal{A}_{exch} + L_{12}\mathcal{A}_{chem} \\ J_{chem} &= L_{12}\mathcal{A}_{exch} + L_{22}\mathcal{A}_{chem} \end{aligned} \tag{10.31}$$

If $L_{12} \neq 0$ and $\mathcal{A}_{chem} \neq 0$, then since at $J_{exch} = 0$

$$\mathcal{A}_{exch} = -RT \ln \Gamma = -\frac{L_{12}}{L_{11}}\mathcal{A}_{chem}$$

the coefficient Γ is different from zero and $L_{11}, L_{22} > 0$. In the case of erythrocytes

$$\Gamma = \exp\left(\frac{L_{12}}{L_{11}}\frac{\mathcal{A}_{chem}}{RT}\right) \approx 220$$

Thus, the operation of the sodium pump is maintained by the coupling of two cycles—the ion-exchange and the chemical cycle.

The phenomenological description shows that the operation of the sodium pump is possible near equilibrium—under linearity conditions.

We shall now turn to a molecular treatment. As has been said above, the source of free energy for active transport is ATP. This compound augments active transport when incorporated into the cell, but it has no effect on the transport when present in the surrounding medium. It was possible to isolate the K^+, Na^+-activated ATPase from cell membranes. This enzyme cleaves ATP only in the presence of K^+ and Na^+ ions. The action of ATP in the membrane is directly associated with active transport—the glycoside ouabain inhibits ATPase at the same concentration at which it interrupts the operation of the sodium pump. The hydrolysis of ATP *in vitro* with the aid of ATPase occurs in two stages. First, ADP is liberated and inorganic phosphate remains bound to the enzyme. This stage is activated by Na^+ ions. The second stage requires K^+ ions and consists in the splitting-off of the phosphate from the enzyme. A similar, though spatial, asymmetry is characteristic of the pump—its activity depends on c^i_{Na} at the internal surface of the membrane and on c^e_K at the external surface. When ATP is cleaved in the membranes, the labeled phosphate is found to have moved from ATP to the phosphoproteins of the membrane. The kinetics of the action of ATPase *in vitro* is described by a sigmoidal plot of the reaction velocity against the concentrations of Na^+, K^+, and ATP. The hydrolysis of one ATP molecule in the membrane is

accompanied by the extrusion of two or three Na^+ ions from the cell.

The K^+, Na^+-activated ATPase is a tetrameric protein with a molecular weight of about 250,000. The protein contains two types of subunits: α (molecular weight 84,000) and β (molecular weight 57,000), two of each of the species. Only the subunit α binds the phosphate label from labeled ATP. On the basis of these facts there has been proposed a model of active transport controlled by the conformational changes of ATPase. A change in conformation induces changes in the affinity of the corresponding active sites for cations. In one conformation the α-subunit binds Na^+ and the β-subunit binds K^+; the situation is reversed in another conformation. The conformational conversion is accompanied by the hydrolysis of ATP; the split-off phosphate is first bound to the α-subunit. Here we again encounter electronic-conformational interactions (see page 212).

Let us now discuss the kinetic model of the process. Suppose that there are two types of active sites capable of binding and exchanging Na^+ and K^+. The first type is nonspecific: these sites take part in passive transport. The sites of the second type belong to the ATPase and provide active transport.

The enzyme (E) catalyzes the hydrolysis of ATP. We denote ATP by the letter X, ADP (and AMP) by Z, and the phosphate by P. The coupling between the ion-exchange reaction and the enzymic reaction is described by the equation

$$Na^i + E^i \cdot K + X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E^i \cdot Na \cdot P + Z + K^i$$

The superscript "*i*" denotes, as before, the internal side and "*e*" the external side of the membrane. On the external side

$$K^e + E^e \cdot Na \cdot P \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E^e \cdot K + P + Na^e$$

The Na·P complex, which breaks down on the outside of the membrane, moves to it from the site of its formation down the concentration gradient. If the affinity of the exchange sites for ions is high and no free sites are available, the movement of the Na·P complex and the exchange for K may be depicted as an exchange reaction:

$$E^i \cdot Na \cdot P + E^e \cdot K \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E^i \cdot K + E^e \cdot Na \cdot P$$

Thus, the generalized force that drives the ions is the chemical potential difference of the product Z which is maintained by the enzymatic reaction.

Suppose further that the exchange of the ions at active sites and the enzymatic reaction are cooperative—the sites interact with one another. A cooperative exchange is observed, for example, in zeolites, where it is governed by the change of the lattice structure due

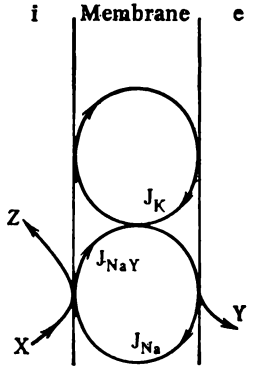


Fig. 10.8. The coupling of ion fluxes in the membrane to a chemical reaction (Y = phosphate).

to the exchange of ions. Accordingly, we introduce the stoichiometric coefficients ρ, ν, and $\varkappa$ different from unity into the reaction equations.

The general scheme of the reactions is (Y ≡ P):

$$\begin{array}{ccc}
\rho Na^i + E^i \cdot K_\nu + \varkappa X & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & E^i \cdot Na_\rho \cdot Y_\varkappa + \nu K^i + \varkappa Z \\
\Big\downarrow\Big\uparrow J_{Na} \quad \Big\downarrow\Big\uparrow k_{-2} & & \Big\downarrow\Big\uparrow k_2 \quad \Big\downarrow\Big\uparrow k_{-2} \quad \Big\downarrow\Big\uparrow J_K \\
\rho Na^e + E^e \cdot K_\nu + \varkappa Y & \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} & E^e \cdot Na_\rho \cdot Y_\varkappa + \nu K^e
\end{array}$$

Here J_{Na} and J_K are passive fluxes. The coupling of the fluxes is shown in Fig. 10.8. The kinetic equations of the reactions are as follows:

$$\begin{aligned}
s^{-1}\dot{c}_X &= -k_1 (c^i_{Na})^\rho (c_X)^\varkappa n^i_K + k_{-1} (c^i_K)^\nu (c_Z)^\varkappa n^i_{Na} \\
s^{-1}\dot{c}^i_{Na} &= -k_1 (c^i_{Na})^\rho (c_X)^\varkappa n^i_K + k_{-1} (c^i_K)^\nu (c_Z)^\varkappa n^i_{Na} + J_{Na} \\
s^{-1}\dot{c}^e_{Na} &= k_2' (c^e_K)^\nu n^e_{Na} - k_{-2} (c^e_{Na})^\rho (c_Y)^\varkappa n^e_K - J_{Na} \qquad (10.32) \\
s^{-1}\dot{c}^i_K &= k_1 (c^i_{Na})^\rho (c_X)^\varkappa n^i_K - k_{-1} (c^i_K)^\nu (c_Z)^\varkappa n^i_{Na} + J_K \\
\dot{n}^i_K &= -k_1 (c^i_{Na})^\rho (c_X)^\varkappa n^i_K + k_{-1} (c^i_K)^\nu (c_Z)^\varkappa n^i_{Na} + s k_3 n^i_{Na} n^e_K + s k_{-3} n^i_K n^e_{Na}
\end{aligned}$$

Here c^i and c^e are the dimensionless concentrations of the ions inside and outside the cell referred to the molar concentrations of water in the corresponding volume, $c^i_{H_2O}$ and $c^e_{H_2O}$; n^i_K and n^e_{Na} are the numbers of the ion-occupied exchange sites of the enzyme per unit area of the internal and external surfaces of the membrane; s is the surface area of the membrane. At the same time,

$$n^i_K + n^i_{Na} = n^i, \quad n^e_K + n^e_{Na} = n^e \tag{10.33}$$

and

$$\begin{aligned} sn^i_K + sn^e_K + c^i_{H_2O} c^i_K + c^e_{H_2O} c^e_K &= n_K \\ sn^i_{Na} + sn^e_{Na} + c^i_{H_2O} c^i_{Na} + c^e_{H_2O} c^e_{Na} &= n_{Na} \end{aligned} \tag{10.34}$$

Relations (10.33) and (10.34) express the retention of the number of exchange sites at the membrane surfaces and of the total number of each species of ions.

In a stationary state the right-hand sides of Eq. (10.32) are equal to zero. In the cell only ATP (X) is used up.

From the first two equations in (10.32) it follows that the active efflux of Na^+ from the cell is determined by the reaction velocity $\dot{c}_X$ and is compensated for in the stationary state by the passive flux J_{Na}.

Assuming the enzymatic reactions to be irreversible, we put $k_{-1} = k_{-2} = k_{-3} = 0$. In this approximation the active efflux of Na^+ from the cell is given by

$$J = k_1 (c^i_{Na})^\rho (c_X)^\varkappa n^i_K = k_2 (c^e_K)^\nu n^e_{Na} = sk_3 n^i_{Na} n^e_K \tag{10.35}$$

Assuming the concentrations of Na^+ and K^+ ions in solutions to be independent of each other, we obtain from the above expression:

$$c^i_{Na} = \left[\frac{k_3 s}{k_1 (c_X)^\varkappa (k_3 s n^i + k_2 (c^e_K)^\nu)} \frac{J (J_0 - J)}{J_\infty - J} \right]^{1/\rho} \tag{10.36}$$

Here

$$J_0 = k_2 n^e (c^e_K)^\nu, \quad J_\infty = sk_3 k_2 n^i n^e (c^e_K)^\nu / (sk_3 n^i + k_2 (c^e_K)^\nu)$$

Thus, c^i_{Na} is defined as a function of the flux at given c^e_K and c_X. Analysis of expression (10.36) shows that the curve of c^i_{Na} vs. J has an inflection point in the interval $0 < J < J_\infty$, which is in agreement with experiment. The same also refers to the curve of c^e_K vs. J. If $\rho = \nu = 1$, i.e., no cooperativity is present, then the curve shows no inflection.

The above-described kinetic model agrees well with experiment and shows that the specific role of the protein-enzyme in membrane transport consists in coupling transport to metabolism.

We are not yet discussing the concrete paths by which the ions are transported across the membrane. The ions are transported through special *channels*, which will be considered in Sections 10.6 and 11.4.

10.6. Transport of Charged Particles Through Membranes

The electric conductivity of cell membranes is about 10^{-3} mho·cm^{-2}; for artificial lipid membranes (see page 351) it is much lower, of the order of 10^{-8} mho·cm^{-2}. These values are many orders of magnitude lower than the electric conductivity of a 0.01M KCl in water, which is equal to 10^{4} mho·cm^{-2}.

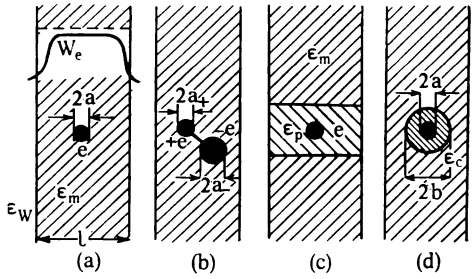


Fig. 10.9. Calculating the energy of an ion in the membrane:
(*a*) The effect of the image forces. (*b*) The formation of ion pairs. (*c*) The hydrophilic pore in the membrane. (*d*) The effect of complex formation.

The low ionic electric conductivity of the lipid membrane, which is treated as a homogeneous medium, is determined by the low dielectric constant (permittivity) of lipids (2-3), this being unfavourable for the incorporation of charged particles. The distribution coefficient of particles between the lipid and the aqueous phase is equal to exp $(-W/RT)$, where W is the energy of the particle in the lipid measured from the energy in water. It is made up of the electrostatic energy and the energy of hydrophobic interaction:

$$W = W_e + W_h \tag{10.37}$$

The first term is the principal one; for spherical particles of radius a it is equal to

$$W_e = \frac{q_0}{a}\left(\frac{1}{\varepsilon} - \frac{1}{\varepsilon_w}\right) \tag{10.38}$$

Here $q_0 = z^2e^2/2kT$ (z is the ionic valence and e is the charge on the electron); ε is the dielectric constant of the lipid; $\varepsilon_w = 81$ is

the dielectric constant of water. At $T = 25°C$, $z = 1$, $q_0 = 28.2$ nm. If $a = 0.2$ nm and $\varepsilon = 3$, then $\exp(-W/RT) = 10^{-20}$.

The energy of the ion in the membrane, however, decreases due to at least four factors (Markin, Pastushenko, and Chizmadzhev, 1977):

1. The membrane has a finite thickness.
2. The ions can form ion pairs inside the membrane.
3. The membrane may have fixed pores (channels) of high dielectric constant.
4. The ion can be transported by an ionophore, a carrier that increases the effective radius a.

These four effects are schematically demonstrated in Fig. 10.9. Let us consider the roles played by these factors.

At the interface between the membrane and the aqueous phase there arise *image forces*. The electrostatic energy W_e of the ion in the membrane decreases somewhat and assumes the shape of the curve shown in Fig. 10.9*a*. Because of this effect the energy is decreased by an amount of the order of a/l, i.e., by a few percent. The barrier height is of the order of a hundred of kJ/mole.

The formation of ion pairs does not give a noticeable gain either—the maximum decrease of the energy is not more than two-fold.

Pores of high polarizability can considerably reduce the energy of the charge in the membrane (Fig. 10.9*c*). At $b \ll l$ the energy of the particle at the pore axis is given by

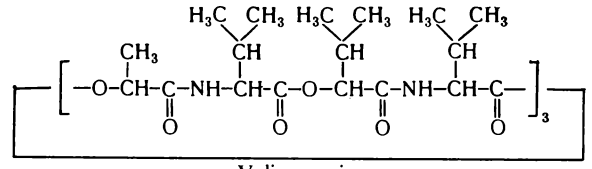
$$W_p = \frac{e^2}{2\varepsilon_p a} + \frac{e^2}{\varepsilon_m b} P\left(\frac{\varepsilon_m}{\varepsilon_p}\right) \tag{10.39}$$

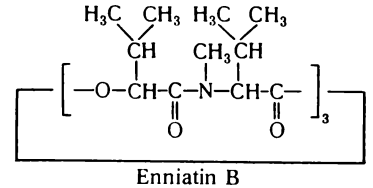
The second term in this formula is due to the image forces acting in the pore walls. The function $P(x)$ has been calculated numerically, its value not exceeding 0.25. If ε_p is comparable with ε_w, then the second term in W_p predominates. At $\varepsilon_m = 2$ the value of $W_p \approx \approx 1180/b$ (nm) kJ/mole.

The fourth factor is *mediated ion transport*. A number of substances serve as *ionophores*—carriers of alkaline cations. Examples are depsipeptides, cyclic antibiotics (say, valinomycin), which have been studied in detail in the works of Shemyakin, Ovchinnikov, Ivanov, and also by Pressman and coworkers. Another group of carriers—monactin, etc.—are macrotetrolides, which are cyclic compounds containing four ether and four ester linkages.

The structure of some cyclic carriers is shown in Fig. 10.10. Ionophores are neutral molecules of high polarizability which form complex spherical complexes with ions (see Fig. 10.9*d* and Fig. 10.11). The energy of the complex in the medium is given by

$$W_c = \frac{e^2}{2\varepsilon_m b} + \frac{e^2}{2\varepsilon_c}\left(\frac{1}{a} - \frac{1}{b}\right) \tag{10.40}$$





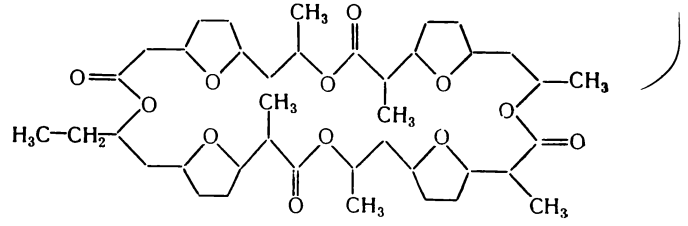


Fig. 10.10. The chemical structure of some cyclic carriers.

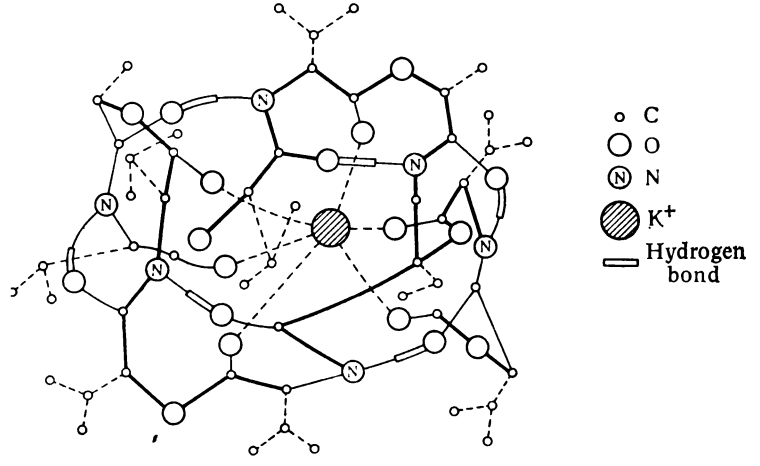


Fig. 10.11. The structure of a complex of valinomycin with K^+.

With a high polarizability of the ionophore the second term may be neglected. At $b \approx 0.5$-1 nm at 25°C

$$W_c = 69.3\text{-}34.4 \text{ kJ/mole}$$

So, the ionophore can considerably lower the barrier to ion penetration. Direct measurements of the temperature dependence of ion trans-

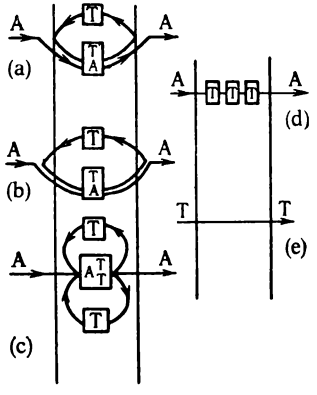


Fig. 10.12. The mechanism of ion transport across the membranes:
(*a*) mobile carriers with a small "merry-go-round" (the carrier T is enclosed in the membrane and complex formation takes place at the membrane solution interface); (*b*) mobile carriers with a large "merry-go-round" (the carrier T is present in the membrane and in the solution; complex formation takes place in the solution); (*c*) cooperative transport (the ion A is ferried by several particles of the carrier T); (*d*) sequential transport; (*e*) direct transport.

port by membranes in the presence of ionophores have demonstrated that the ion transport obeys the Arrhenius equation

$$g = A \exp(-H^{\ddagger}/RT) \qquad (10.41)$$

The enthalpy of activation $H^{\ddagger}$ is equal to 136.5 kJ/mole for monactin, to 231 kJ/mole for valinomycin, and to only 39 kJ/mole for gramicidin-A. Monactin and valinomycin are mobile carriers, while gramicidin-A forms a polar pore in the membrane. In the presence of valinomycin and monactin the conductivity sharply decreases when the lipid membrane is being "frozen"; no such effect is observed in the presence of gramicidin-A. The transport facilitated by the formation of pores is a sequential process—a pore can be formed by several successively arranged molecules between which the transfer of an ion takes place. The ion can be transported not by one but simultaneously by several ionophore molecules—this is a cooperative mechanism. Various mechanisms of induced ion transport are shown schematically in Fig. 10.12.

As has been pointed out (see page 354), membrane transport may be associated with the presence of kinks—mobile structural defects in the liquid-crystalline lipid phase. The possibilities of the inter-

pretation of this kind of effects on the basis of the theory of *dislocations* have not yet been completely exhausted.

While discussing the membrane transport of univalent cations, we have closely approached the problems pertaining to excitable membranes, to the generation and propagation of the nervous impulse.

10.7. Molecular Reception

The reception of external signals by the sense organs involves receptor cells and cell systems which are specific to each sense organ and in which the signal (visual, hearing, smell, taste, or touch) is transformed to nervous impulses. The nature of such transformations is still obscure in many respects. Here we shall deal with the reception of molecular signals, which include smell and taste. The visual reception will be considered in Chapter 14 and the mechanical reception in Sec.12.6.

Various receptor cells have fine, hairlike projections—antennae, whose plasma membranes contain proteins specific to a particular type of reception. These antennae may consist of so-called microvilli and stereocilia (the latter are formed upon differentiation of microvilli)—cilia and flagella and their derivatives. These also include fibrous proteins. Flagella have a "$9 \times 2 + 2$" arrangement, i.e., a bundle of 11 thin fibres, consisting of a circle of nine double fibres around two central single microtubules (see Sec. 12.6). In other cases, flagella have a "$9 \times 2 + 0$" structural pattern.

Smell reception is very important to a number of organisms of vertebrates and invertebrates. Insects use the language of odours—the special chemical substances secreted externally by insects, called *pheromones*, serve as messages for the other members of their own species. There exist triggering pheromones used by ants as an alarm signal or as a trail indicating the path to food. Butterflies release pheromones as sex attractants, a fact discovered by Fabre. A number of pheromones have been obtained in a pure form, in particular, *bombykol*—the sex attractant of the female silkworm moth *Bombyx mori*. Butenandt and his associates extracted 4 mg of bombykol from 313,000 butterflies and determined its structure:

$$H_3C{-}CH_2{-}CH_2{-}CH{=}CH{-}CH{=}CH{-}(CH_2)_7{-}CH_2OH$$

The sensitivity of butterflies to bombykol is very high—for the male to be excited 10^{-18} g in 1 cm^3 of solvent (i.e., 2500 molecules) will suffice. The statements occasionally appearing in the literature that insects communicate by electromagnetic signaling have no grounds. Orientation by odour has been established for many organisms—mammals, fish, etc.

The high sensitivity of smell receptors indicates that odours are transmitted by molecules. The threshold concentrations of odorous substances sensed by man are 4×10^{-7} for skatole, 4.4×10^{-8} for ethylmercaptan, and 5×10^{-9} mg/litre for trinitrobutyltoluene. An odorous substance must be sufficiently volatile and capable of dissolving in water and in lipids—the receptor cells are embedded in a mucous, aqueous environment and the substance must pass through the membranes.

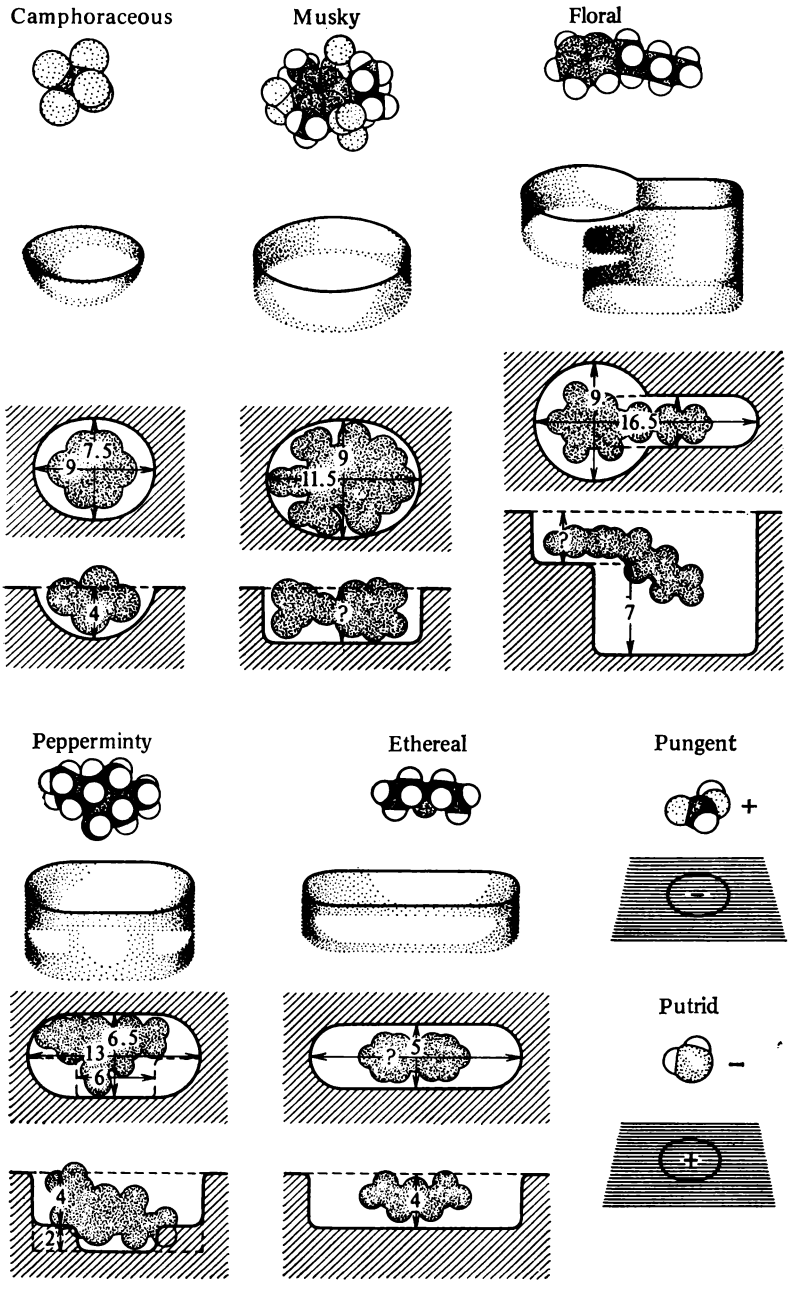
There has been made an assumption that smell reception is based on the resonance of atomic oscillations of the molecules of an odorous substance and of the molecular structures of the receptor. This theory, known as the vibrational or quantum theory, cannot be considered to have been theoretically substantiated. The nature of the smell and its intensity correlate poorly with the vibrational spectra of the substance. Substances that have strongly differing spectra often have similar odours and vice versa. The vibrational theory contradicts the elementary physical considerations. The nasal cavity is practically a closed system: it is, so to speak, a sort of black body, and if molecules get into it, their radiation emission must be in equilibrium with the cavity walls. The vibrations in such a system cannot be perceived.

Moncrieff (1951) suggested that odorous molecules produce their effects by fitting closely into the receptor sites of the cells. Based on this idea, Amoore made an extensive search, looking for clues in the chemical characteristics of odorous compounds and came to the conclusion that there were seven primary odours, and in 1952 his findings were summed up in a stereochemical theory of olfaction that identified the seven odours (see Table 10.3).

Comparison of the structures of substances that have such odours has shown that the smell is determined not by chemical composition but by the shape and size of their molecules. Figure 10.13 shows the

Table 10.3. The Seven Primary Odours with Chemical and More Familiar Examples (identified by J.E. Amoore, J.W. Johnston, Jr., and M. Rubin)

Primary odour	Chemical example	Familiar substance
1. Camphoraceous	Camphor	Moth repellent
2. Musky	Pentadecanolactone	Angelica root oil
3. Floral	Phenylethyl methyl ethyl carbinol	Roses
4. Pepperminty	Menthone	Mint candy
5. Ethereal (ether-like)	Ethylene dichloride	Dry-cleaning fluid
6. Pungent	Formic acid	Vinegar
7. Putrid	Butyl mercaptan	Bad egg


Camphoraceous
Musky
Floral
7.5
9
4
9
11.5
?
9
16.5
?
7
Peppermint
Ethereal
Pungent
+
-
Putrid
6.5
13
6
4
2
?
5
4
+
-

molecular structures and shapes of the corresponding receptor sites. The pungent and putrid odours are determined not by the structure but by the charge—electrophilic substances with small molecules have a pungent odour (HCOOH, SO_2, Cl_2), whereas nucleophilic substances have a putrid odour (H_2S). Mixed odours made up of two or more primaries arise if the different groups of the same substance are located at different receptor sites. On the basis of the Amoore receptor-site hypothesis, there has been possible to carry out the directed synthesis of an odorous substance. Thus, there is every ground for believing that smell reception is based on the direct recognition of molecular structure effected by means of weak interactions (cf. page 58). No better olfaction theory has yet been worked out.

Fesenko and his colleagues (1978) studied the tissue membranes of the frog and mouse olfactory epithelia and found that they contain structures having high affinity for camphor. Their specific sensitivity towards this odorous compound, the low dissociation constant, and their absence from other tissues point to the presence in the membrane of olfactory receptor molecules. A number of facts indicate that the olfactory receptor for camphor is a membrane protein with a molecular weight of about 125,000. Such a protein must have a molecular diameter of the order of 8-10 nm. This value is consistent with the size of particles detected with the aid of an electron microscope on the flagella of the olfactory receptor cells of a calf and a mouse.

The manner by which the binding of the molecules of odorous compounds to the membranes of smell receptor cells triggers a nerve impulse is unknown.

Taste reception is molecular and not structural in character. The sour (acid) taste is due to the presence of hydrogen ions and the salty taste, to the presence of anions, such as Cl^-. The bitter and sweet sensations arise when the receptor cells are stimulated by substances of widely differing structures (quinine and sodium sulphate are bitter, sugar and saccharin are sweet). There have been discovered specific proteins of vegetable origin which exhibit high taste activity. Two of these, monellin and thaumatin, which have an intense sweet taste, are chemostimulants. On the other hand, the glycoproteid mi-

Fig. 10.13. Olfactory receptor sites are shown for each of the seven primary odours, together with molecules representative of each odour. The shapes of the first five sites are shown in perspective and (with the molecules silhouetted in them) from above and the side. The molecules are (left to right) hexachloroethane, xylene musk, alpha-amylpyridine, *l*-menthol, and diethyl ether. Pungent (formic acid) and putrid (hydrogen sulphide) molecules fit because of charge, not shape.

raculin is a taste modifier—after its action on the tongue the acid is perceived as a sweet substance. Presumably, miraculin becomes bound to the plasma membrane. The acid alters its conformation, stimulating the "sweet" portion of the membrane. The sensation of taste must in the final analysis boil down to molecular recognition.

The membrane receptor molecules—proteins, glycoproteins (glycoproteids), etc.—participate in the most important biological phenomena. The generation of immunity is described in Sections 4.8 and 17.8. The intercellular interactions responsible for morphogenesis (see Sec. 17.7) are effected via molecular, chemical signals. This is proved by direct experiments, in which the interaction of cells was suppressed by inserting a piece of cellophane between them. When cellophane was replaced by agar-agar, the interaction was restored.

It has long been shown that the separated cells of a sponge again unite when the sponge is placed in sea water, well-developed small sponges being formed as a result. Regeneration is found to be species-specific. Evidently, recognition, which leads to the ordering of cells, requires molecular signaling, and the contact and adhesion of cell surfaces.

It has been found that the permeability of cell membranes is increased upon contact. This contact is preceded by signaling via special substances secreted by cells. In some cases it proved possible to identify the secreted substances. They turned out to be glycoproteins having a sedimentation constant of 62.5S and a spherical shape 80 nm in diameter; they have extensions 110 nm in length and 4.5 nm in thickness. In other cases, signaling is effected via small molecules, in particular, by cAMP (see pages 44 and 45).

The movement of cells, as a result of which they are sorted out, is best accounted for by their differential adhesion. Edelman (1967) proposed a hypothesis of the nature of mutual recognition of cells, their movement and also of cell growth based on surface modulation. According to this hypothesis, the surfaces of cells of a particular species contain molecules of specific glycoproteins. These are cleaved by certain proteases, and the fragments left at the surfaces recognize each other. Thus, the state of the surface is modulated by proteases, which thereby control the adhesion of cells in developing tissues. These glycoproteins are responsible for the association of cells.

We may note the similarity between various phenomena of chemical modulation of the surfaces: antibody-antigen interactions; injection of viruses; interactions with hormones; the fusion of a sperm with an egg, etc. The Edelman hypothesis is important and may serve as the basis for further investigations.

Chapter 11 Physics of the Nerve Impulse

11.1. The Axon and the Nerve Impulse

Nervous excitation is propagated via nerve fibres called axons. The nervous system of higher organisms is commonly subdivided into the central nervous system (CNS) and the peripheral nervous system (PNS). The latter is made up of axons which transmit signals and ganglia of the vegetative nervous system. Axons are communications for afferent (or sensory) neurons, which carry signals from receptors (the sense organs) to the central nervous system, and for efferent (or motor) neurons which transmit signals from the CNS to the effectors (muscles). Axons are long, slender extensions or processes of centrally located nerve cells.

The study of the nervous impulse is the traditional problem in biophysics. The speed at which the nerve impulse is conducted was originally measured by Helmholtz. In 1902 Bernstein worked out the membrane theory of excitation. The investigations carried out by Hodgkin, Katz, Huxley, Tasaki, and other scientists have disclosed the principal mechanisms of generation and propagation of the nerve impulse.

Thus, much is known about the functioning of the axon. But the present-day status of science permits one to model only formally the functioning of the central nervous system, and we are still far from understanding its higher functions—memory and thinking.

Figure 11.1 is a schematic representation of the structure of the nerve cell—the *neuron*. The cell receives messages from many other neurons via their thin extensions which form junctions called *synapses* between the cell body and its short outgrowths—*dendrites*. The length of axons in the body of large animals may amount to several metres.

The contacts with other cells are formed not only in synapses. The larger part of the neuron surface is wrapped by *glial* or *Schwann cells*. Their role is not clear yet. The membranes of Schwann cells form a *myelin sheath* of myelinated axons shown schematically in Fig. 11.1. This sheath is interrupted at regular intervals (every 1-2 mm of the length of the axon) by constrictions or the *nodes of*

Ranvier. These nodes are about 1 μm long. In the region of the nodes of Ranvier the axon membrane comes into contact with its environment. There also exist unmyelinated axons.

The myelin sheath of the axon consists of a multiple helical wrapping of the cell membrane of Schwann cells around the axon. The myelin is thus a multilayer (up to 250 layers) membrane which protects the axon membrane against the environment. Figure 11.2

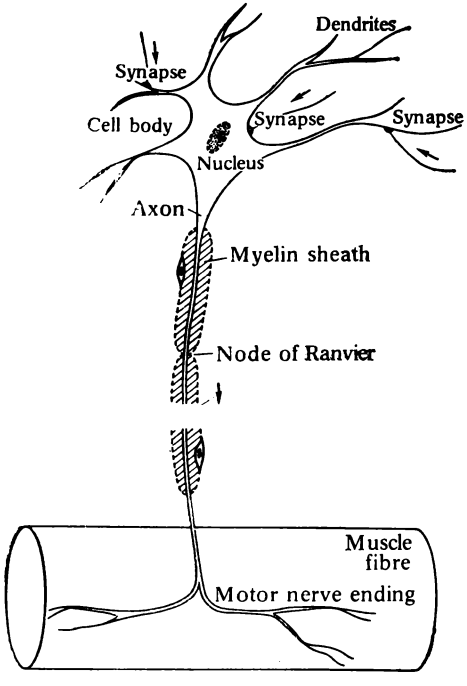


Fig. 11.1. Schematic representation of the structure of the nerve cell (the motor neuron).

shows the diagram of the myelin membrane structure obtained by means of the X-ray diffraction technique.

It has long been established that the nerves and muscles are capable of generating electromotive forces called *biopotentials*. The electrical activity of cells manifests itself in the form of short discharges, each of which lasts about 1 ms. Modern experimental instruments make it possible to amplify and record these signals. It has been found that the nerve activity is always accompanied by electrical phenomena. Figure 11.3 shows the impulses in the optic nerve of the crab *Limulus* set up by a light flash of 1 s (Hartline, 1934). The frequency and character of a sequence of impulses depend on the

intensity and spectral composition of the light. The magnitude and duration of an impulse are independent of the nature and amount of excitation.

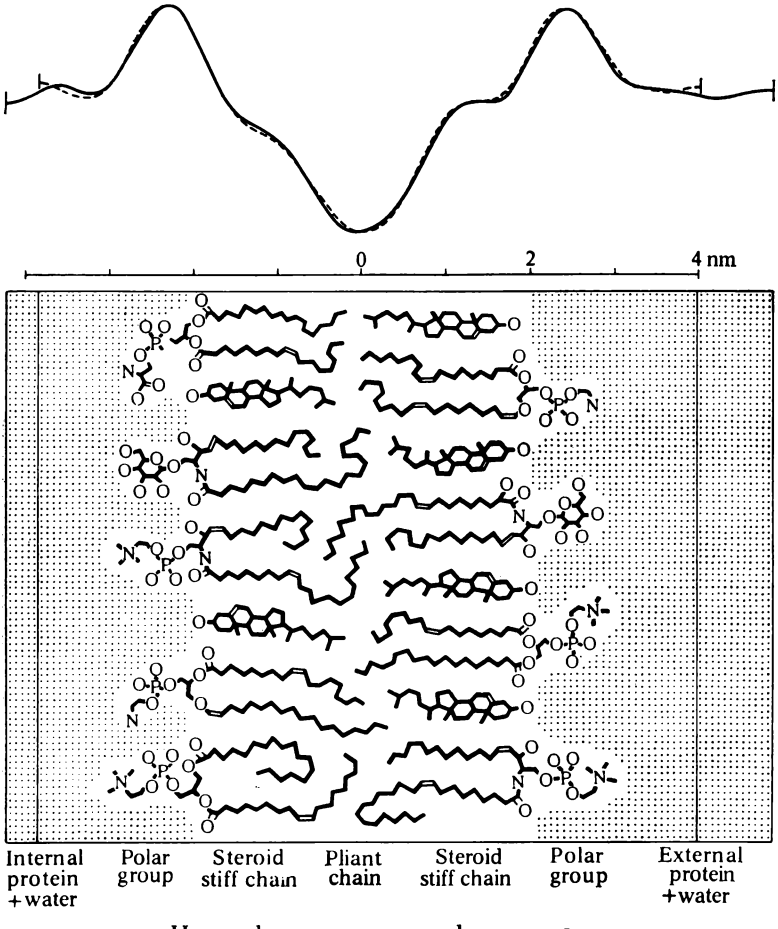


Fig. 11.2. Schematic representation of the myelin membrane structure.
Top: electron-density profiles for the optic nerve (the dotted line) and the sciatic nerve (the solid line).

An electric signal corresponding to an individual impulse conducted along the axon is called the *action potential* or *spike potential.* This is the basic unit of information transmitted along the nerve fibre.

The rate of travel of the nerve impulse, v, is 1-100 m/s; it is faster for myelinated than for unmyelinated axons. For instance, for the myelinated fibres of a cat v = 10-100 m/s, and for the giant axon of a squid (the fibre diameter 600 μm!) $v \approx 25$ m/s; for the unmyelinated fibres of a cat v = 0.7-2.3 m/s.

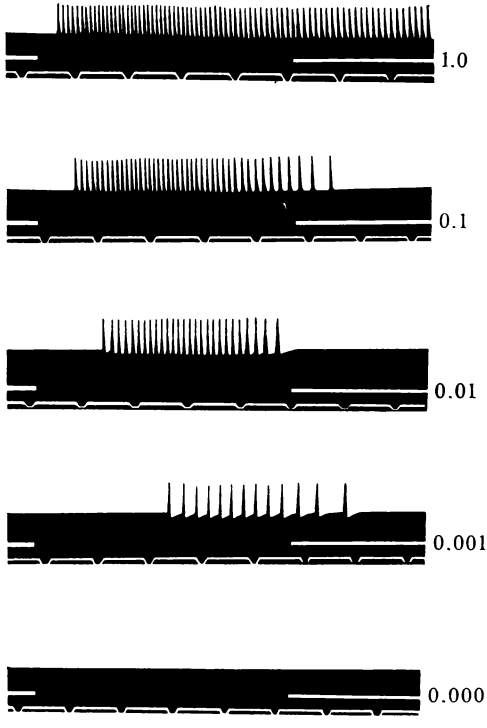


Fig. 11.3. The impulses in the optic nerve fibre of *Limulus* stimulated by a light flash of 1 s. The numbers on the right indicate the relative light intensity. The illumination time is indicated by a break in the upper white line; the time interval on the lower white line is 0.2 s.

Despite the fact that the *axoplasm* is a solution of electrolytes, it does not serve as an impulse conductor. The specific resistance (resistivity) of the axoplasm ranges from 10 to 100 ohms·cm. The resistance per unit length of a fibre with a diameter of 1 μm is 10^9-10^{10} ohm·cm, which exceeds the resistance of a copper wire of the same diameter by 10^8 times. In such a conductor the losses and leakage are high. But the nerve impulse is transmitted by the axon to distances of up to several metres without being damped and distorted.

An important role in the biophysics of the nerve impulse has been played by methods of investigation with isolated axons—the insertion of microelectrodes into the axon and perfusion, i.e., the squeez-

ing-out of the axoplasm from the fibre and its replacement by artificial solutions. Work with squid giant axons is especially convenient. A simple experiment consists in introducing two microelectrodes into the axon: one serves for electric stimulation, and the second is used for measuring the generated potential (Fig. 11.4).

Experiment shows that the magnitude and time course of action currents are independent of the magnitude of the stimulating current

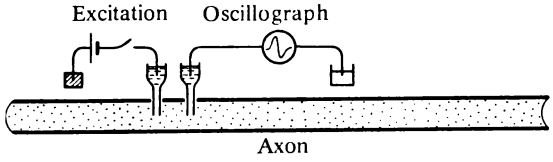


Fig. 11.4. A diagram for investigating the nerve impulse with the aid of two microcells.

and that an action potential does not arise if the electric stimulus does not reach a certain threshold value. The nerve fibre acts according to the all-or-none law—an action potential of constant value either arises or does not arise at all.

In order to excite the axon, a certain minimum quantity of electricity is required. As the duration Δt of the impulse transmitted is reduced the current I must be increased. For prolonged rectangular impulses there exists a minimum current sufficient for excitation; a weaker current is inefficient at any duration. Figure 11.5 is a plot of the threshold current I_{th} against Δt; the curve has the shape of a hyperbola. It is usually described by the following empirical formula:

$$I_{th} = \frac{a}{\Delta t} + b \tag{11.1}$$

The quantity b is called the *rheobase*; it is the minimum value of the exciting current at $\Delta t \to \infty$. With very short impulses, i.e., at small Δt, the current $I_{th} \approx a/\Delta t$, i.e., the quantity $I_{th}\Delta t \approx a$ is a constant that characterizes the threshold level of the amount of electricity. From formula (11.1) it follows that at $I_{th} = 2b$ the quantity $\Delta t_{ch} = a/b$ (*chronaxy*). In reality, however, at $\Delta t = a/b$ the current $I_{th} \approx 1.4b$—formula (11.1) does not hold for the entire range of Δt values.

If two successive stimuli separated by a certain time interval are applied to the nerve fibre, the behaviour of the fibre will depend on this time interval. Immediately after a nerve impulse has been initiated, the given part of the fibre is in the absolute *refractory state*,

i.e., it cannot be excited again. This is followed by a relative refractory state in which the threshold potential is somewhat increased. The duration of the entire refractory period varies from one to a few milliseconds.

In the resting state the axoplasm is charged negatively with respect to its environment. The resting potential is about —80 mV. The threshold excitation has a potential of about —50 mV, while the action potential is equal to 40 mV. This is schematically shown in Fig. 11.6.

The transmission of a nerve impulse is determined by a change in the state of the axon membrane upon its excitation. As pointed

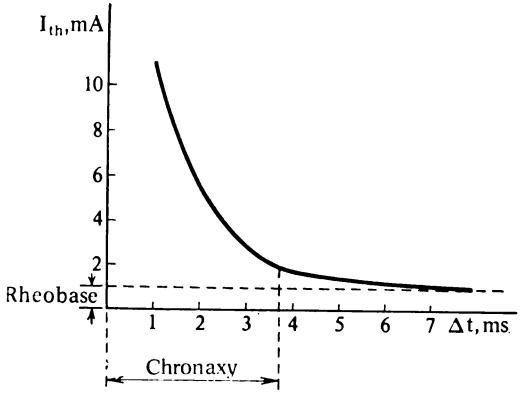


Fig. 11.5. The dependence of the threshold current on the duration of the impulse.

out above, in the resting state of the cell membrane, as a result of the active transport of ions, the concentration of K^+ ions in the cytoplasm (axoplasm) by far exceeds their concentration in the environment; for Na^+ ions the situation is reversed. Table 11.1 gives the composition of the axoplasm of the axons of *Loligo*, the composition of blood plasma and the related composition of seawater.

When an exciting potential is applied to the membrane, first its permeability to Na^+ ions is increased. The Na^+ ions enter the axon, as a result of which the inner surface of the membrane changes the sign of its charge from negative to positive. In other words, the membrane undergoes *depolarization.*

If we make use of formula (10.23) to express the potential difference

$$\varphi \equiv \Delta\psi = \frac{RT}{\mathscr{F}} \ln \frac{P_K c_K^i + P_{Na} c_{Na}^i + P_{Cl} c_{Cl}^e}{P_K c_K^e + P_{Na} c_{Na}^e + P_{Cl} c_{Cl}^i} \qquad (11.2)$$

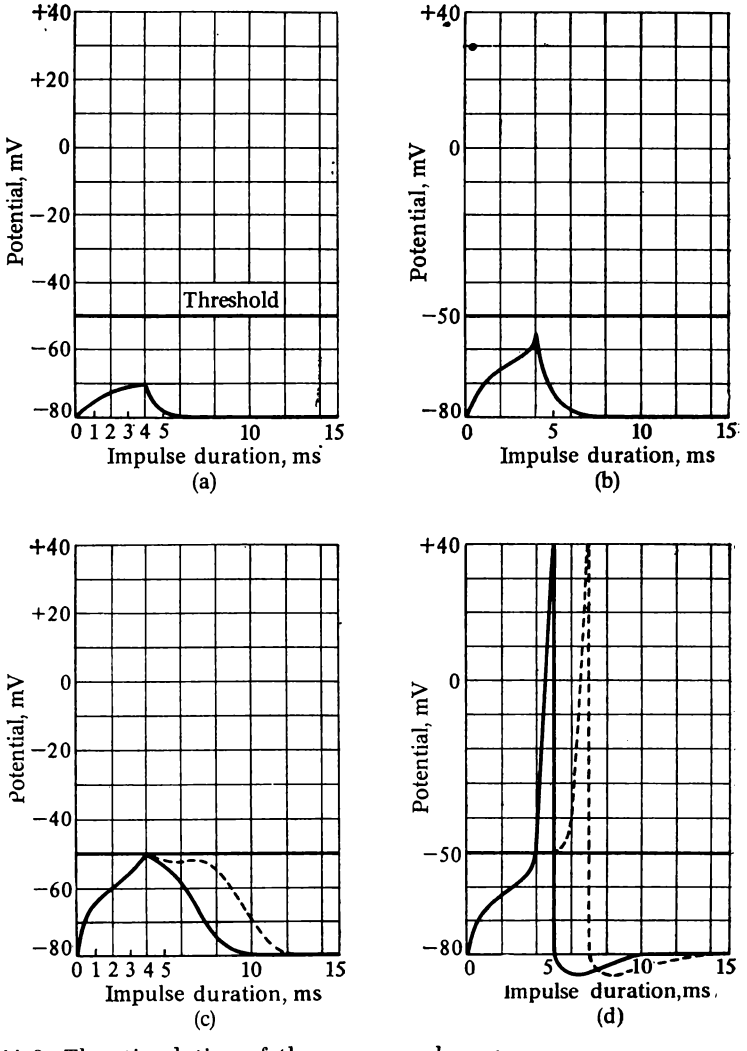


Fig. 11.6. The stimulation of the axon membrane:
(*a*) and (*b*) subthreshold impulses; (*c*) the threshold attenuated impulse; (*d*) the appearance of an action potential with a superthreshold impulse; the dotted lines indicate the moving impulse.

Table 11.1. The Concentration of Ions and Other Substances in the Axon of the Squid *Loligo*

Substance	Concentration, mmole per kg H_2O			Substance	Concentration, mmole per kg H_2O		
	Axoplasm	Blood	Seawater		Axoplasm	Blood	Seawater
K^+	400	20	10	Succinic and fumaric acids	17	—	—
Na^+	50	440	460	Orthophosphate	2.5-9.0	—	—
Cl^-	40-150	560	540	ATP	0.7-1.7	—	—
Ca^{2+}	0.4	10	10	Arginine phosphate	1.8-5.7	—	—
Mg^{2+}	10	54	53	Water	865	870	966
Isethionic acid	250	—	—				
Aspartic acid	75	—	—				
Glutamic acid	12	—	—				

then the resting state is characterized by the ratio of the ion permeabilities

$$P_K : P_{Na} : P_{Cl} = 1 : 0.04 : 0.45$$

In an excited state corresponding to the generation of an action potential

$$P_K : P_{Na} : P_{Cl} = 1 : 20 : 0.45$$

Let us emphasize that the interpretation of excitation as the result of the alteration of the ionic permeabilities is phenomenological and does not disclose the molecular mechanism of the process.

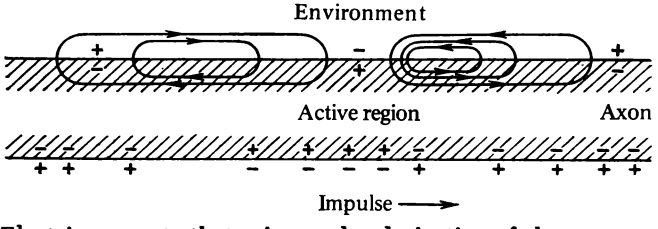


Fig. 11.7. Electric currents that arise on depolarization of the axon membrane.

The depolarization of the membrane gives rise to currents which are short-circuited through the external conducting medium. These currents stimulate the next portion of the axon (Fig. 11.7).

Thus, according to the membrane theory, when the impulse is generated, a certain channel is opened through which Na^+ ions get into the interior of the axon, causing the membrane to be depolariz-

ed. When the impulse is being generated, the sodium channel is closed and the potassium channel is opened. The ejection of K^+ ions leads to the restoration of the negative charge on the inner surface of the membrane. In this way the membrane is *repolarized.* After the

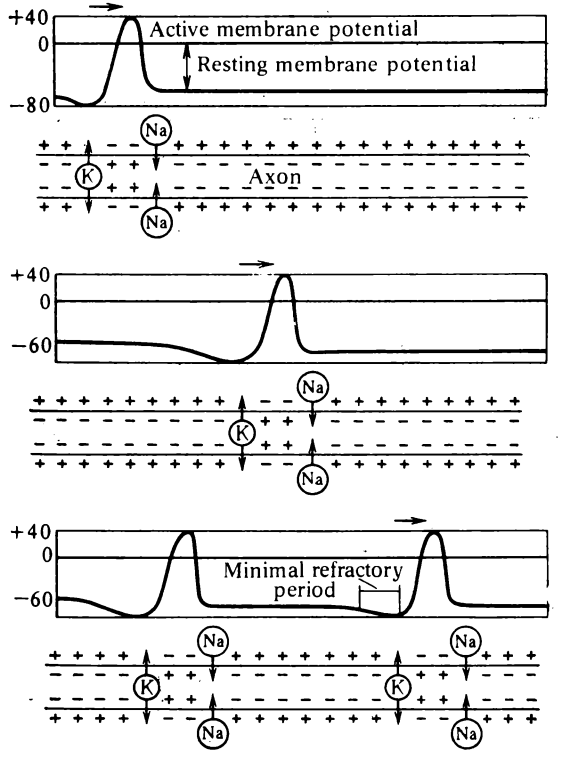


Fig. 11.8. The propagation of the nerve impulse is accompanied by a change in the permeability of the axon membrane to ions.

refractory period has expired, the membrane becomes capable of transmitting a new impulse (Fig. 11.8).

As can be seen from Fig. 11.7, the net lateral current flowing across the cross-section of the axon and the environment is equal to zero—at all points the internal currents are equal in magnitude and opposite in direction to the external currents. But the density of the longitudinal current and the longitudinal potential difference between two points in the interior of the axon are different from those in the exterior of the axon. The axon membrane has a resistance of

1000 ohms·cm^2 and a capacitance of 1 μF/cm^2, which corresponds to a lipid bilayer having a thickness of 5 nm, a dielectric constant of $\varepsilon = 5$ and a specific resistance of 2×10^9 ohms·cm. During the generation of an impulse the conductance of the membrane increases by about 10^3 times. The electrical properties of the membrane can be modelled by an equivalent circuit shown in Fig. 11.9. This figure shows only one element of the membrane, and we should visualize

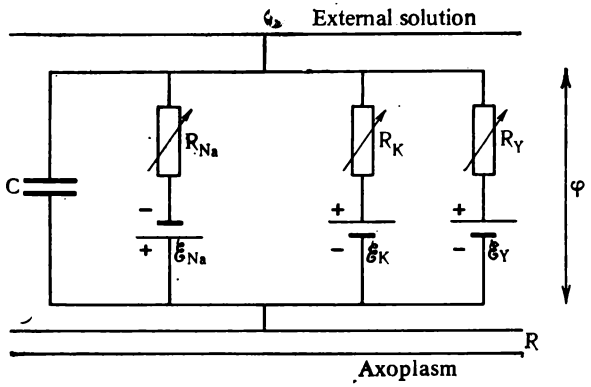


Fig. 11.9. An equivalent diagram of an element of the excitable membrane of the nerve firbre.

a long linear sequence of such elements, forming a continuous cable. The resistance R characterizes the axoplasm; the external solution is present in a large excess and is depicted by a conductor without resistance. The sodium and potassium "batteries" $\mathcal{E}_{Na}$ and $\mathcal{E}_K$ determine the generation of an impulse, and the additional "battery" $\mathcal{E}_Y$ depicts the movement of other ions which does not change upon excitation.

Thus, the propagation of an impulse is a self-sustained process; the nerve impulse resembles a burning spark traveling along the length of a fuse.

As has been said above, the nerve impulse travels in a myelinated fibre faster than in an unmyelinated fibre. In the myelinated axon the transmission of the impulse occurs in a *saltatory* way: the nerve impulse jumps from one node of Ranvier to another. Potassium and sodium channels open and close only at the nodes of Ranvier—the myelinated portions are well insulated. The myelin sheath has a low capacitance, which accounts for the high rate of transmission of the nerve impulse. The saltatory conduction of the nerve impulse is shown schematically in Fig. 11.10.

The myelination inherent in the axons of invertebrates provided their evolutionary advantage, since a myelinated fibre may be 25 times thinner than an unmyelinated fibre at the same rate of transmission of the impulse.

The membrane theory is directly supported by experiments involving the perfusion of axons. It has been possible to squeeze the axoplasm out of the giant axon of *Loligo* without the membrane being

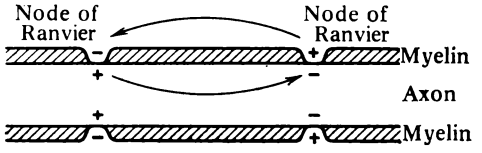


Fig. 11.10. Schematic representation of the saltatory transmission of the nerve impulse.

damaged and to replace the axoplasm with an artificial solution. The perfused axons are capable of conducting up to 10^5-10^6 impulses for several hours.

The resting potential disappears when the concentration of K^+ ions on the inner side of the membrane becomes equal to that on the outer side. When KCl in the axoplasm is replaced by NaCl, the resting potential falls down to zero. At the same time, the resting potential is only slightly sensitive to concentrations of K^+ ions less than 20 mM. Perfusion experiments show that the resting potential is really controlled by K^+ ions. When KCl is replaced by NaCl and the concentration of K^+ ions is considerably increased (up to 600 mM), the situation is reversed and the internal part of the fibre is found to be electrically positive (and not negative) relative to its surrounding medium. Indeed, when the fibre immersed in an isotonic solution of KCl is filled with an analogous solution of NaCl, the inner solution is found to be positively charged relative to the external solution and φ is 50-60 mV.

Direct experiments with labeled atoms show that the conduction of the nerve impulse is associated with the increase of the velocity of movement of potassium and sodium along the concentration gradients. In the giant axon of the cuttlefish *Sepia* with each nerve impulse there is observed an influx of Na^+ ions equal to 10.3×10^{-12} mole/cm^2 and an efflux of the same ions equal to 6.6×10^{-12} mole/cm^2. The net influx of Na^+ (3.7×10^{-12} mole/cm^2) per impulse is approximately equal to the influx of K^+ ions. Indeed, for the voltage of the capacitor having a capacitance of 1 μF to be changed by 120 mV there is required a charge of 0.12×10^{-6} C,

which is equivalent to 1.3×10^{-12} mole/cm^2 of a univalent cation. Thus, the measured influx of Na^+ is more than sufficient for the action potential to be set up.

With each nerve impulse in an unmyelinated axon about 20,000 Na^+ ions pass across 1 μm^2 of the surface area. In a myelinated axon the number of Na^+ ions that enter the axon per impulse through every node of Ranvier is 6×10^6. The surface area of the membrane at the node is 20 μm^2, which means that 300,000 Na^+ ions pass across every 1 μm^2. The density of the ion current at the node is about 10 times higher than in unmyelinated giant axons.

The generation of a nerve impulse is accompanied by thermal phenomena—the nerve is first heated and then cooled. It has been shown that this effect depends on three factors. First, when the action potential is generated, there occurs an energy dissipation caused by the ion currents. Second, this is accompanied by an energy dissipation brought about by a change in the separated charges on the two sides of the membrane—a decrease in the transmembrane potential corresponds to the heat released. This contribution, however, is an order of magnitude smaller than the ionic contribution. Finally, the opening of the ion channel is accompanied by the liberation of heat, which is then absorbed. In general, theory is in agreement with experiment, according to which the heat released by the rabbit vagus is 100 $\mu J/g$ per impulse and the heat absorbed 93 $\mu J/g$.

11.2. Propagation of the Nerve Impulse

We know relatively simple electrochemical processes that are similar to the movement of the nerve impulse. For example, such a process occurs when a pure iron wire is immersed in concentrated nitric acid. The wire surface becomes coated with a passivating oxide film. If the film is destroyed at some spot, the iron will vigorously dissolve. This process spreads along the wire, the activation front advances, this being followed by the repassivation of the iron. This is the model thoroughly studied by Lillie and also by Markin. Theory shows that there is a far-reaching formal analogy between the Lillie model and the mode of propagation of the nerve impulse. Other models have also been developed.

Nerve excitation begins with a local generation of the action potential. The impulse then propagates along the nerve axons and *syncytia*, which are systems composed of numerous fibres, the impulse passing from one fibre to another. The syncytial structure is characteristic of smooth muscles, which coat the hollow organs of animals. Nerve excitation is transmitted from one neuron to another by means of synaptic transmission, the nature of which is a special problem.

Before we consider the generation of the nerve impulse determined by physico-chemical processes in an active membrane, let us turn

to a physical treatment of the propagation of the nerve impulse. Especially important in this field are the works of Hodgkin and Huxley which are devoted to the solution of the entire problem of theoretical calculations of the shape of the action potential and the speed at which the nerve impulse is transmitted.

The basic equation for the membrane current that arises in a homogeneous portion in response to a shift in potential may be written in the form

$$I = C\frac{d\varphi}{dt} + I_i \tag{11.3}$$

where φ is the potential difference between the two sides of the axon membrane; $C\dot{\varphi}$ is the capacitive current associated with a change in the density of ions on the external and internal surfaces of the membrane; I_i is the ionic current determined by the movement of charged particles across the membrane. For the squid axon I_i is composed of the currents of Na^+ and K^+ ions (I_{Na} and I_K) and the leak current I_Y. Hodgkin and Huxley have demonstrated experimentally that these ionic currents are linearly dependent on the corresponding electrochemical potentials:

$$I_{Na} = g_{Na}(\varphi - \varphi_{Na})$$

etc. The Hodgkin-Huxley theory is based on the separate measurement of the potassium and sodium currents. To do this, the *voltage-clamp* technique is used. Let us make use of the model shown in Fig. 11.9. By short-circuiting the membrane with a metallic wire, we can sharply lower the membrane potential down to zero. The capacitor C is discharged, after which a current is created only by ions that pass through R_{Na} and R_K. Upon sharp depolarization of the membrane to a potential value between 20 and 110 mV the net ionic current is first the inward Na^+ current along the concentration gradient. After the lapse of about 1 ms there arises an outward K^+ current. If the total potential difference is φ_{Na}, there is observed only a potassium current. At $\varphi > \varphi_{Na}$ the sodium current is directed outward. The resolution of the net current into its components can be carried out by changing the concentrations of Na^+ and K^+ ions. Figure 11.11 shows the observed curves of I vs. t. If the axon is immersed in seawater, the net current $I = I_{Na} + I_K$ will be described by curve *1*. When Na^+ is replaced by choline, there is observed a net potassium current (curve *2*). The difference between these two curves (curve *3*) yields I_{Na}. When the potential is suddenly displaced inside the fibre by +56 mV (a "short-circuiting" of the membrane) the sodium conductance g_{Na} first increases rapidly from zero to 25×10^{-3} ohm^{-1}·cm^{-2} and then declines (Fig. 11.12). The potassium conductance increases slowly, following the course of an S-shaped curve, and after 5-6 ms it reaches a constant level. Upon repolariza-

tion of the membrane g_{Na} drops considerably faster than does g_K. With various shifts of the membrane potential there is observed a regular change in the course of the $g_{Na}(t)$ and $g_K(t)$ curves. The potassium and sodium conductances of the axon membrane depend significantly on the presence of Ca^{2+} ions (see Sec. 11.3).

Hodgkin and Huxley proposed a satisfactory description of the observed changes based on a special model. It is presumed that K^+ can pass through the membrane only if four single-charged particles

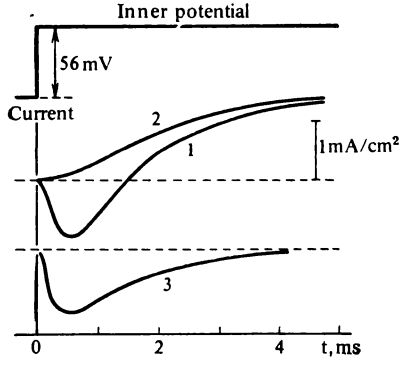


Fig. 11.11. Resolution of the membrane current I into the potassium and sodium components:
1. $I = I_{Na} + I_K$; *2.* Sodium is replaced by choline, $I = I_K$; *3* = 1-2, $I = I_{Na}$. The deviation upward is the outflowing current.

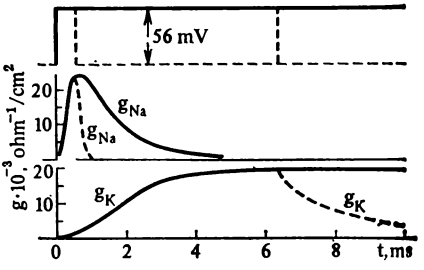


Fig. 11.12. The variation of g_{Na} and g_K during the depolarization of the membrane by 56 mV. The solid lines indicate the conductance on prolonged depolarization; the dotted lines indicate the conductance on repolarization of the membrane in 0.6 and 6.3 ms.

approach a certain region of the membrane under the influence of an electric field. The potassium conductance is equal to

$$g_K = \bar{g}_K n^4 \tag{11.4}$$

where $\bar{g}_K$ is the maximum conductance; $n < 1$ is the probability of the approach of the particle. The kinetic equation for n has the form

$$\dot{n} = \alpha_n (1 - n) - \beta_n n \tag{11.5}$$

where the rate constants α_n and β_n depend, at constant concentration of Ca^{2+} ions, only on the membrane potential φ. When the positive potential inside the fibre increases, α_n increases and β_n decreases.

Since the potassium channel may be considered to be wide open upon approach of a negatively charged particle, the quantity n may

be treated as the probability of two Ca^{2+} ions being expelled from the potassium channel.

The next assumption consists in that the sodium channel is opened if three activating particles find their way into the given portion and one blocking particle is simultaneously expelled. Denoting the probability of the entry of an activating particle by m and the probability of the exclusion of the blocking particle by h, we obtain:

$$g_{Na} = \bar{g}_{Na} m^3 h \tag{11.6}$$

For m and h we write equations analogous to Eq. (11.5):

$$\dot{m} = \alpha_m (1 - m) - \beta_m m \tag{11.7}$$

$$\dot{h} = \alpha_h (1 - h) - \beta_h h \tag{11.8}$$

When the interior of the axon becomes more positive, the values of α_m and β_h increase and those of α_h and β_m decrease.

At a fixed voltage all the values of α and β are constant; hence, n, m, and h are exponentially dependent on time. Calculation of the curves of g_K and g_{Na} versus t and φ by means of these formulas gives good agreement with experiment.

The density of the net membrane current is given by the followingt expression [cf. Eq. (11.3)]:

$$I = C \frac{\partial \varphi}{\partial t} + (\varphi - \varphi_K) \bar{g}_K n^4 + (\varphi - \varphi_{Na}) \bar{g}_{Na} m^3 h + (\varphi - \varphi_Y) \bar{g}_Y \tag{11.9}$$

where C is the membrane capacity per unit surface area. The firs term in Eq. (11.9) is the capacitive current and $\bar{g}_Y$ is the conductance for Cl^- and other ions which corresponds to the leak.

Let us initiate the action potential by a short pulse. After the excitation is completed both the radial and the longitudinal current in a given section are equal to zero. Hence, $I = 0$ for $t > 0$. The dependence $\varphi(t)$ is found from Eq. (11.9) if the initial value of φ is taken as the boundary condition. Theory gives good agreement with experiment (Fig. 11.13).

Let us consider a certain portion of the axon having a length l. The radius of the axoplasm is a and its resistance is R_a. In a stationary state the current that flows into the portion under consideration is equal to zero. We have

$$\frac{\pi a^2}{R_a} \left[-\frac{\partial \varphi}{\partial x} + \left(\frac{\partial \varphi}{\partial x} + \frac{\partial^2 \varphi}{dx^2} l \right) \right] = 2\pi a l I \tag{11.10}$$

where I is given by Eq. (11.9). The left-hand side of Eq. (11.10) contains a divergence of the longitudinal current, and the right-hand side contains the membrane current. If the nerve impulse trav-

els at a constant speed v, then the following wave equation is valid:

$$\frac{\partial^2\varphi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\varphi}{\partial t^2} \qquad (11.11)$$

and Eqs. (11.9), (11.10), and (11.11) yield the *Hodgkin-Huxley equation*:

$$\frac{a}{2Rv^2}\frac{d^2\varphi}{dt^2} = C\frac{d\varphi}{dt} + (\varphi - \varphi_K)\bar{g}_K n^4 + (\varphi - \varphi_{Na})\bar{g}_{Na} m^3 h + (\varphi - \varphi_Y)\bar{g}_Y \qquad (11.12)$$

A numerical solution of Eq. (11.12) that corresponds to the finite values of φ gives the rate of propagation v which agrees well with the experimental value. For example, for the squid axon the speed calculated with the aid of the experimental values of a, R_a, C, φ_K, φ_{Na}, φ_Y, $\bar{g}_K$, $\bar{g}_{Na}$, and $\bar{g}_Y$ is equal to 18.8 m/s. Experiment gives 21.2 m/s. At this value of v the potential after completion of the spike returns to the resting level.

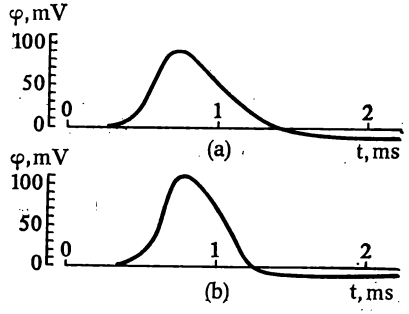


Fig. 11.13. The propagating action potentials:
(*a*) theory; (*b*) experiment (the squid giant axon at 18.5°C).

As a result of subsequent investigations, it became possible to resolve and simplify the problem. In calculating the rate of propagation of the impulse one may distract from its precise shape and consider the movement of the impulse along an electric cable formed by the axon membrane and characterized by certain values of resistance and capacitance. The inductance of the fibre plays no significant role. The approaches along this line are realized, in particular, in the works of Kompaneyets (1966, 1971).

Four phases can be distinguished in the propagation of the nerve impulse. First, the potential difference increases from its resting value φ_0 to the threshold level φ_1. Second, it increases from φ_1 to φ_{max}, up to the action potential. Then, the system returns to the initial value φ_0 and, finally, passes to a refractory state. The speed of propagation of a single impulse is determined only by the first two phases, in which the potential increases. By specifying an initial rectangular current pulse and taking into account the values of capacitance and resistance per unit length of the axon and its radius, we can calculate the velocity of propagation of the impulse both in unmyelinated and myelinated axons in reasonable agreement with experi-

ment. The resulting dependence of the velocity of propagation on the parameters indicated follows also from dimension considerations.

The theory of the propagation of the nerve impulse was further developed in the works of Markin, Chizmadzhev, *et al.* (1969-1974).

The propagation of excitation along an unmyelinated homogeneous fibre is described by the general equation

$$\frac{\partial \varphi(x, t)}{\partial t} = \frac{1}{RC} \frac{\partial^2 \varphi(x, t)}{\partial x^2} - \frac{I}{C} \tag{11.13}$$

where φ is the membrane potential measured from the resting membrane potential; R is the sum of the external and internal resistances; C is the capacitance of the unit fibre length; I is the ionic current

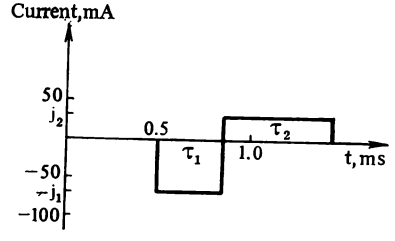


Fig. 11.14. An approximation of the ionic current for stimulation of the membrane.

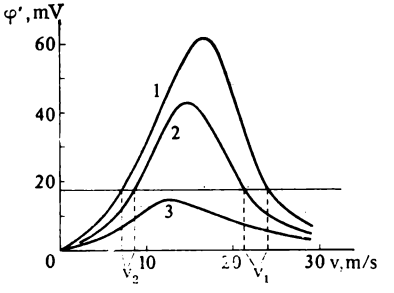


Fig. 11.15. The rate of travel of the nerve impulse at various leakage conductivities g_{leak}:
1—0 mho·cm^{-2}; *2*—1 mho·cm^{-2}; *3*—5.74 mho·cm^{-2}.

flowing across the membrane (which is also referred to unit length of the fibre); $I > 0$ if the current flows outward.

Once the threshold is reached the ionic current flows into the interior of the fibre and after a certain time has elapsed it changes its direction and flows outward. The ionic current may be approximated by two rectangular "tables" (Fig. 11.14). In this approximation Eq. (11.13) is solved without much difficulty. The coordinate $\xi = x - vt$ is introduced, where v is the velocity of propagation of the impulse. Equation (11.13) is rewritten in the form

$$\frac{d^2\varphi(\xi)}{d\xi^2} + vRC \frac{d\varphi(\xi)}{d\xi} - RI(\xi) = 0 \tag{11.14}$$

We find wave solutions on the condition $\varphi(\infty) = 0$ and with finite φ values at $\xi \to -\infty$. Equation (11.14) is satisfied by two solutions with different values of v. Figure 11.15 shows the curves of the stim-

ulating threshold φ' vs. v for three values of the leak conductance $\bar{g}_Y$. The following values of the parameters are assumed for the squid axon (see Fig. 11.14): $j_1 = 63$ μA, $j_2 = 40$ μA, $\tau_1 = 36$ ms, $\tau_2 = 0.55$ ms ($j_1\tau_1 = j_2\tau_2$), $C = 0.157$ μF/cm; the specific resistance of the axoplasm $\rho = 50$ ohms·cm, the axon diameter $a = 0.05$ cm; $\varphi' = 18.5$ mV. The straight line with $\varphi' = 18.5$ mV crosses the φ' (v) curve at two points at sufficiently low values of $\bar{g}_Y$. The high and low velocities (v_1 and v_2, respectively) are approximately given by

$$v_1 \approx \frac{j_1/(\varphi' RC^2) - 2/(r_m RC)}{[j_1/(\varphi' RC^2) - 1/(r_m RC)]^{1/2}} \tag{11.15}$$

$$v_2 \approx \left[\frac{2\varphi'}{Rj_1\tau_1(\tau_1+\tau_2)}\right]^{1/2} \tag{11.16}$$

Here r_m is the resistance of the membrane per unit axon length in the resting state. Since $j_1 \propto a$, $R \propto a^{-2}$, and $C \propto a$, we find that $v \propto a^{1/2}$, which is in agreement with experiment.

The solution of v_1 is stable and that of v_2 is unstable. Suppose that the velocity v_1 has increased by a small amount. Then, as seen from Fig. 11.15, the potential at the midpoint of the impulse declines down to a value below the threshold level φ'. As a result, the impulse is slowed down. But if v_1 decreases, the potential at the front point will exceed the threshold level and the impulse will be accelerated. For v_2 the inverse relations are valid.

With account taken of the experimental values of the parameters for the squid axon ($r_m = 6.37 \times 10^3$ ohms·cm) we obtain $v_1 = 21$ m/s in excellent agreement with experiment.

Theory thus shows that the velocity of propagation of the impulse is determined by the electrical and geometrical parameters of the axon, almost independently of the shape of the original impulse. Theory also gives solutions for fibres of variable cross-section and for branching fibres. Markin, Chizmadzhev, *et al.* have developed the theory of interaction of nerve fibres in fascicles and stems, the theory of propagation of excitation in syncytia.

11.3. Generation of the Nerve Impulse

The current-voltage characteristic of the axon membrane is specific. It is not only nonlinear but contains a portion with a negative slope (Fig. 11.16*a*). This is a consequence of the dependence of Na and K conductances on the membrane potential.

Let us consider the generation of the action potential in a homogeneous section of the axon. Since at the first step of depolarization the K conductance is very low, we assume the ionic current to be the sum of I_{Na} and the leak current I_Y. The leak current is defined

by Ohm's law as the difference between the active potential ψ and the resting potential ψ_0 divided by the resistance of the membrane:

$$I_Y = \frac{\psi - \psi_0}{R} \equiv \frac{\varphi}{R}$$

The time dependence of I_{Na} may be approximated by a rectangular well if we assume that a current of constant value I_{Na} exists within the time interval from the moment t^* its peak value is reached to the moment t_3 when the Na-current falls off. During the remaining time interval $I_{Na} = 0$.

Experiments carried out at a fixed potential show that the dependence of I_{Na} on φ is nonlinear, i.e., the dependence of g_{Na} on φ is nonlinear. However, with an instantaneous jump of the potential the conductance g_{Na} has no time to change and, according to the experiments carried out by Hodgkin and Huxley, I_{Na} is linearly dependent on $\varphi - \varphi_{Na}$:

$$I_{Na} = g_{Na} (\varphi - \varphi_{Na}) \tag{11.17}$$

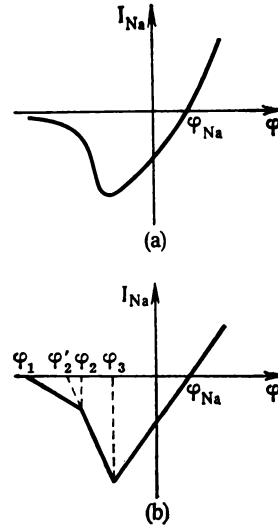


Fig. 11.16. An experimental current-voltage curve (*a*) and its approximation by a broken line (*b*).

Let us replace the experimental curve of I_{Na} vs. φ shown in Fig. 11.16*a* by a broken line with discrete absolute values of conductance (Fig. 11.16*b*):

$$\left| \frac{dI_{Na}}{d\varphi} \right| = \begin{cases} g_1 \text{ at } \varphi_1 < \varphi < \varphi_2 \\ g_2 \text{ at } \varphi > \varphi_2 \end{cases} \tag{11.18}$$

The voltage-current characteristic of the peak Na-current is described by the following formulas:

$$I_{Na} = \begin{cases} 0 & \varphi < \varphi_1 \\ -g_1(\varphi - \varphi_1) & \varphi_1 < \varphi < \varphi_2 \\ -g_2(\varphi - \varphi_2') & \varphi_2' < \varphi < \varphi_3 \\ g_2(\varphi - \varphi_{Na}) & \varphi > \varphi_3 \end{cases} \tag{11.19}$$

with $g_1 < g < g_2$, where $g = R^{-1}$. With the membrane being excited by a direct current from a source of voltage $\tilde{\varphi}$ we have $I = g\tilde{\varphi}$. Let us consider the response of the membrane to such excitation.

If $\tilde{\varphi} < \varphi_1$, then $I_{Na} = 0$. The equation for the membrane current $I_m = C\partial\varphi/\partial t + I_Y$ assumes the form

$$C\frac{d\varphi}{dt} = g(\tilde{\varphi} + \varphi_0 - \varphi) \tag{11.20}$$

Its solution

$$\varphi - \varphi_0 = \tilde{\varphi}[1 - \exp(-gt/C)] \tag{11.21}$$

describes the passive behaviour of the membrane with stimuli of threshold intensity. The magnitude of the stimulus $\tilde{\varphi} = RI$ sets up the limit for the increase of φ. If at time $t = t_0$ the stimulating current is switched off, then

$$C\frac{d\varphi}{dt} = g(\varphi_0 - \varphi) \tag{11.22}$$

with the following initial condition:

$$\varphi(t_0) = \varphi_0 + \tilde{\varphi}[1 - \exp(-gt/C)]$$

The solution gives the rapidly falling-off voltage:

$$\varphi - \varphi_0 = \tilde{\varphi}[\exp(gt_0/C) - 1]\exp(-gt/C) \tag{11.23}$$

If the stimulating current is sufficiently high, then at time $t = t_1$ the growing potential will reach the value φ_1. At $t \geqslant t_1$

$$C\frac{d\varphi}{dt} = g\tilde{\varphi} - g(\varphi - \varphi_0) + g_1(\varphi - \varphi_1) \tag{11.24}$$

The solution of Eq. (11.24) has the form

$$\varphi - \varphi_0 = \tilde{\varphi} + \frac{\tilde{\varphi} - (\varphi_1 - \varphi_0)}{g - g_1}\left\{g_1 - g\exp\left[-\frac{g - g_1}{C}(t - t_1)\right]\right\} \tag{11.25}$$

It describes the exponential increase of $\varphi - \varphi_0$ up to the value

$$\varphi_\infty - \varphi_0 = \tilde{\varphi} + \frac{g_1}{g - g_1}[\tilde{\varphi} - (\varphi_1 - \varphi_0)] \tag{11.26}$$

The local response determined by the difference between the potential shift and the corresponding passive response is found by subtracting Eq. (11.21) from Eq. (11.25). The local response appears to have a sigmoidal shape, which agrees with experiment (Fig. 11.17).

Using relation (11.21) for $t = t_1$

$$\varphi_1 - \varphi_0 = \tilde{\varphi}[1 - \exp(-gt_1/C)]$$

we eliminate $\exp(-gt_1/C)$ from Eq. (11.25) and obtain

$$\varphi - \varphi_0 = \tilde{\varphi} + \frac{\tilde{\varphi} - (\varphi_1 - \varphi_0)}{g - g_1}\left[g_1 - g\exp\left(-\frac{g - g_1}{C}t\right)\left(\frac{\tilde{\varphi}}{\tilde{\varphi} - \varphi_1 + \varphi_0}\right)^{\frac{g-g_1}{g}}\right] \tag{11.27}$$

The dependence of φ on the stimulus $\tilde{\varphi}$ is nonlinear. The curve of φ vs. $\tilde{\varphi}$ shown in Fig. 11.18 coincides with the experimental curve.

We see that the semi-empirical model under consideration reveals

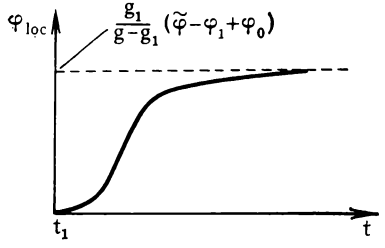


Fig. 11.17. The dependence of the local response on time.

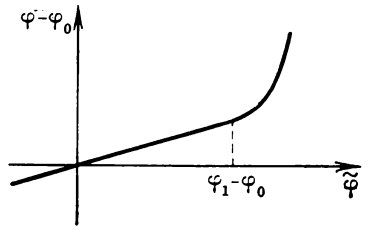


Fig. 11.18. The dependence of the shift of the membrane potential on the applied stimulus.

the basic features of the local response: the S-shaped dependence on time, the nonlinear dependence on the magnitude of the stimulus. These features are due to the appearance of a current flowing across the membrane and determined by the negative differential conductance, whose absolute value g_1 is lower than the membrane conductance g.

If the limit $\varphi_\infty - \varphi_0$ to which the potential shift tends exceeds φ_2, the system is found to be in the region where the Na-current is determined by the negative differential conductance, g_2 being greater than g. Since $g > g_1$, we may assume that $g_2 \gg g_1$. The change of the membrane potential is described by the following equation:

$$C\frac{d\varphi}{dt} = g\tilde{\varphi} - g(\varphi - \varphi_0) + g_2(\varphi - \varphi_2) \tag{11.28}$$

When the stimulus is applied for a prolonged period of time, the moment t_2 this equation becomes valid is reached when the potential begins to rise, this portion of the curve being described by Eq. (11.25). The solution of Eq. (11.28) has the form

$$\varphi = -\frac{g}{g_2 - g}(\tilde{\varphi} + \varphi_0) + \frac{g_2}{g_2 - g}\varphi_2 + \\ + \left[\varphi(t_2) + \frac{g}{g_2 - g}(\tilde{\varphi} + \varphi_0) - \frac{g_2}{g_2 - g}\varphi_2\right]\exp\left[\frac{g_2 - g}{C}(t - t_2)\right] \tag{11.29}$$

The exponential is positive; φ increases or decreases, depending on the sign of the pre-exponential term. If by the time t_2 the potential becomes equal to φ_2, then the pre-exponential term is positive on the condition $\tilde{\varphi} > \varphi_2 - \varphi_0$. The potential rises up to the time $t = t_3$, when, as a result of the membrane being inactivated, the

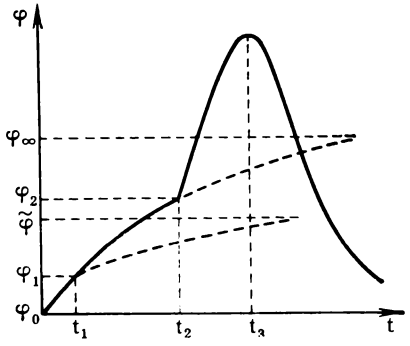


Fig. 11.19. The growth of the membrane potential with time on prolonged stimulation by a current of a superthreshold level:

t_1—the start of the local response; t_2—the moment of switching on the I_{Na}-current; t_3—the moment of switching off the I_{Na}-current.

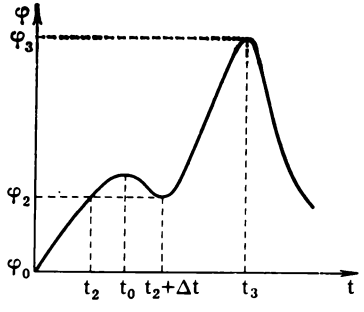


Fig. 11.20. The growth of the membrane potential with time on stimulation by a direct-current impulse.

Na-current falls off. If by this time the stimulus has ceased to act, the behaviour of the system is described by the equation

$$C \frac{d\varphi}{dt} = -g(\varphi - \varphi_0)$$

and its solution is

$$\varphi - \varphi_0 = [\varphi(t_3) - \varphi_0] \exp\left[-\frac{g}{C}(t - t_3)\right]$$

The potential falls off exponentially down to the level of the resting potential φ_0. The entire course of the dependence of φ on t is described by the curve in Fig. 11.19.

If the stimulus acts for a short time interval t_0, the situation is different. Suppose the action of the stimulus is stopped as early as the isotonic step of the build-up of the membrane potential. By the time t_2 the potential $\varphi = \varphi_2$ but the Na-current is not switched on until a certain time interval Δt has elapsed. But during this time interval the stimulus ceases to act and the potential begins to fall.

If by the time $t = t_2 + \Delta t$ it has not become less than φ_2, i.e.,

$$\varphi(t_2+\Delta t) = \varphi_0 + \tilde{\varphi}\left[\exp\left(\frac{g}{C}t_0\right) - 1\right]\exp\left[-\frac{g}{C}(t_2+\Delta t)\right] \geqslant \varphi_2 \tag{11.30}$$

the Na-current is switched on, a spike appears, and the time course of the potential follows the curve shown in Fig. 11.20.

Using Eq. (11.30), we can obtain the relationship between the current I and the duration of its action which is in quantitative agreement with experiment.

It has been shown on the basis of the adopted model that the development of the spike potential on the axon membrane begins from the moment the membrane potential reaches a value at which the conductance is negative and exceeds in absolute magnitude the membrane conductance.

11.4. Ionic Channels

The above-described models of the generation of the impulse are phenomenological and do not disclose the molecular mechanisms involved.

But how are the ions transported across the axon membrane (cf. Sec. 10.6)? Are they transported by means of special carriers or through channels in the membrane? Experiment solves this alternative in favour of the channels. In the absence of permeant Na^+ and K^+ ions the high-frequency conductance of the membrane is equal to zero. If some mobile charged groups functioned in the membrane, there must have been observed an electric current. It is possible to estimate the quantitative characteristics of the channels, the number of which per 1 μm^2 of the membrane surface is several hundreds. The conductance of a single sodium channel is 4×10^{-12} mho, and that of a single potassium channel $12 \times \times 10^{-12}$ mho. The conductance (permeability) of the first channel is about 10^7 ions per second, and that of the latter, about 10^6 ions/s. Nonetheless, it is difficult to say what the channel is in the structural sense. An ionic channel may be looked upon as a sort of "vectorial enzyme" which catalyzes the transport of ions. The activity of such an "enzyme", i.e., a lipoprotein complex experiencing conformational rearrangements (kinks, see page 354), is regulated by an electric field.

The ion exchange via channels is a source of energy for the propagation of the nerve impulse. The driving force for this process is provided by the sodium pump which operates at the expense of the energy of ATP (Sec. 10.5).

The ionic channels are highly selective. It is, however, possible to make the Na^+-channels transport other cations by blocking the

K^+ channels by tetraethylammonium. The same can be done with K^+ channels. Table 11.2 presents the relative permeabilities of channels to various cations. This table shows the selectivity of the channels. The cross-section of Na^+ channels determined from these data is about 0.3×0.5 nm^2. There are observed large differences in the permeabilities to ions of similar sizes. The hydroxylamine, $H_3N^+—OH$, and hydrazine, $H_3N^+—NH_2$, pass through the Na^+ channel and methylamine, $H_3N^+—CH_3$, does not pass at all. Hille presumes that in the narrowest part of the channel there are oxygen atoms, one of which is charged negatively. The cations with OH and NH_2 groups slip through the channel, forming hydrogen bonds with O^-; the CH_3 group is not capable of doing this. The ions move through the channel one after another.

Table 11.2. The Relative Permeabilities of Na^+ and K^+ Channels at Nodes of Ranvier to Monovalent Cations (Hille, 1972)

Ion	Crystallographic radius of ion, nm	P_{ion}/P_{Na} in Na-channel	P_{ion}/P_K in K-channel
Lithium	0.120	0.93	0.018
Sodium	0.190	1.00	0.010
Potassium	0.266	0.09	1.00
Thallium	0.280	0.33	0.300
Rubidium	0.296	—	0.910
Ammonium	0.300	0.16	0.130
Hydroxylamine	0.330	0.94	0.025
Hydrazine	0.333	0.59	0.029
Cesium	0.338	—	0.077

The sodium channel is tightly sealed by one molecule of tetrodotoxin (the toxic principle of the Japanese fish fugu, *Spheroides porphyreus*) or saxitoxin (the toxic principle of *Gonyaulax*, the dinoflagellate of "red tides"). Accordingly, these poisons paralyze nerve conduction. The use of these substances has made it possible to determine the number of channels per unit surface area of the membrane and to study the gating current (see below).

Hodgkin and Huxley have established the exponential relationship between the initial change of the membrane potential φ and the peak value of g_{Na}. To the change of φ by 4-6 mV to the side of depolarization there corresponds an e-fold increase in g_{Na}. One might think that this change is the result of the action of an electric field on the disposition of some ions in the channel. Considering that the conductance g_{Na} is proportional to the fraction of these ions displaced by the field, we obtain

$$g_{Na} = \frac{g_{Na,\,max}}{1 + b \exp(-z\alpha\varphi\mathcal{F}/RT)} \tag{11.31}$$

where z is the valency of the hypothetical ion; α is the number which expresses the fraction of voltage φ that acts on the ion; b is a factor which is independent of φ. The average experimental value of $z\alpha$ is equal to 3.5 according to available data. Hodgkin and Huxley presumed that such hypothetical ions determine the activation and inactivation of Na^+ and K^+ channels. Admittedly, these are Ca^{2+} ions. Indeed, the change in the concentration of Ca^{2+} ions has a strong effect on the excitability of nerve and muscle cells. A theoretical model has been worked out which accounts for the shape of the curve of g_{Na} vs. φ and the shift of this curve caused by the change of the concentration of Ca^{2+} ions, which block the channels. It is supposed that the membrane contains activating particles capable of being in two states, A and B, the transitions $A \rightleftharpoons B$ being associated with the performance of work in an electric field. Such transitions may consist of the movement of a charged molecule and the rotations of dipoles or the change of the conformation of the macromolecule. In state A the particle forms an $A \cdot Ca$ complex with Ca^{2+}, and in state B it contributes to the passage of Na^+ ions through the channel. The model corresponds to the following scheme:

$$A \cdot Ca \rightleftharpoons Ca^{2+} + A \rightleftharpoons Ca^{2+} + B$$

The solution of the corresponding kinetic equations yields:

$$\ln[Ca^{2+}] - \frac{z\alpha\varphi\mathcal{F}}{RT} = \text{const} \qquad (11.32)$$

and the quantity $\Delta\varphi$ with an e-fold change of the Ca^{2+} concentration is equal to $RT/(z\alpha\mathcal{F})$, which gives 7 mV at $z\alpha = 3.5$ in good agreement with experiment.

On the basis of similar conceptions it is also possible to consider the inactivation of the membrane.

Experiment shows that when the conductance of the axon membrane undergoes change, i.e., upon its excitation, there really occur some structural changes. There have been observed changes in light scattering (Cohen, Hille, and Keynes), in fluorescence (Tasaki *et al.*), in birefringence of the membrane (Berestovsky), caused by the change of φ. The properties of individual channels have been studied by Ermishkin.

Direct proofs for the displacement of charged particles that enter the membrane structure under the action of the field have been obtained by Armstrong and Bezanilla.

The structural rearrangement of the channel caused by a change in the external electric field manifests itself in the presence of an additional component of the displacement current, called the *gating current*. The measurements were carried out on axons in which the internal medium was replaced by a potassium-free solution and

the sodium ions were expelled from the external solution. The conductance of the channels was suppressed by tetrodotoxin. In order to average large fluctuations, use was made of a storage element. The experiment was as follows: a positive rectangular pulse was applied to a fibre that had been preliminarily hyperpolarized to −90 or −100 mV. At the start and finish of the pulse there were observed transient currents directed to the exterior and interior of the axon. For the capacity charging current to be taken into account, analogous measurements were carried out for a negative voltage pulse. The results of the first and second experiments were summed up algebraically. In this way there was singled out the small asymmetric component of the gating current. This current depends exponentially on time, its relaxation time being close to the relaxation time of the m-variable in the Hodgkin-Huxley model (page 390). One may therefore expect that the gating current is really associated with particles that open and close the Na^+ channel. Other facts have also been obtained which confirm this supposition.

11.5. Synaptic Transmission

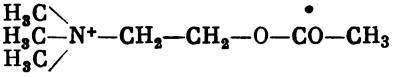
When the impulse reaches a nerve fibre ending, it either performs its function at a neuromuscular (myoneural) junction, inducing muscular contraction, or is transmitted to another nerve fibre. In either case there is accomplished a synaptic transmission of the nervous impulse. The synapse is a functional intermembrane junction between two excitable cells.

Modern biology has at its disposal a wealth of information concerning the synaptic transmission. These data are still insufficient for a molecular theory to be constructed but allow us to outline paths for its development.

The distance between the presynaptic and the postsynaptic membrane—the *synaptic cleft*—varies from 15 to 20 nm. In the myoneural junction the gap is even larger, up to 50-100 nm. At the same time, there are synapses with closely spaced and even merging presynaptic and postsynaptic membranes. Accordingly, two modes of synaptic transmission (or synaptic conduction) occur: electrical and chemical. Chemical transmission is accomplished if the cleft is large; with cells being in close contact direct electrical transmission is possible. We shall here be concerned with chemical transmission.

In chemical transmission, the nerve impulse that reaches a nerve fibre terminal involves a chemical mechanism, which amplifies the electrical signal. This mechanism consists of the release of a special substance, a *synaptic transmitter substance*, which is synthesized and stored in nerve endings, its reception by specific centres of the

postsynaptic membrane, and of the alteration of the membrane permeability, as a result of which a new impulse is triggered. Among the transmitter substances are, in the first place, acetylcholine (ACh)



and other related compounds. Acetylcholine is concentrated in tiny membrane-bond sacs—the vesicles of presynaptic endings. The presynaptic mechanism is neurosecretion, i.e., the release of acetylcholine from the vesicles under the action of the nerve impulse. Estimates of the number of acetylcholine molecules released per impulse give several millions.

The postsynaptic membrane of the myoneural junction has been found to contain a high concentration of *acetylcholine esterase* (AChE), an enzyme that catalyzes the hydrolysis of acetylcholine. It has been shown that the receptor substance is a special hydrophobic protein. This protein was isolated from the membranes of nerve endings. It has a high affinity for acetylcholine and other cholinergic substances. De Robertis proposed a model of the postsynaptic membrane (Fig. 11.21). According to De Robertis, the membrane has discrete receptor regions. The protein receptor, which has a hydrophobic surface, passes through the lipid layers of the membrane and creates channels in it. The walls of the channels are formed by four parallel protein molecules. Acetylcholine esterase is present as an individual molecule. Acetylcholine attaches itself to the active site of the protein receptor on its external surface. As a result of this binding (a strong ionic interaction), conformational and translational transformations occur in the membrane, leading to changes in its ionic permeability.

The role of acetylcholine esterase (AChE) consists mainly in hydrolyzing acetylcholine. If acetylcholine were not hydrolyzed, it would have blocked AChE, and synaptic transmission would have been stopped. Such action is exhibited by the competitive inhibitors of AChE, in particular, the vegetable poison curare and other alkaloids. Consequently, muscle innervation is also blocked.

It has been shown experimentally that the transmitter ACh does really increase the permeability of the end plate of the neuromuscular junction to Na^+ and K^+ ions simultaneously and to the same extent. The permeability to Cl^- is not affected by ACh. The electrical effect of the action of the transmitter is similar to a short-time "puncture" of the membrane.

Let us return to the presynaptic region. Data are available, which indicate that the release of small amounts of ACh occurs also in the resting state, independently of the nerve impulse. The impulse strongly increases this activity during the time of the order of

1 ms. Spontaneous neurosecretion is not caused by the leakage of ACh from nerve endings by way of accidental diffusion. Spontaneous neurosecretion is a quantized release of acetylcholine, in which portions of ACh are released at random moments of time by the "all-or-none" law from the discrete points of the terminal axon membrane. The magnitude of a "quantum" of ACh does not depend on the changes of the membrane associated with the nerve impulse. The impulse alters by hundreds of thousands of times the probability

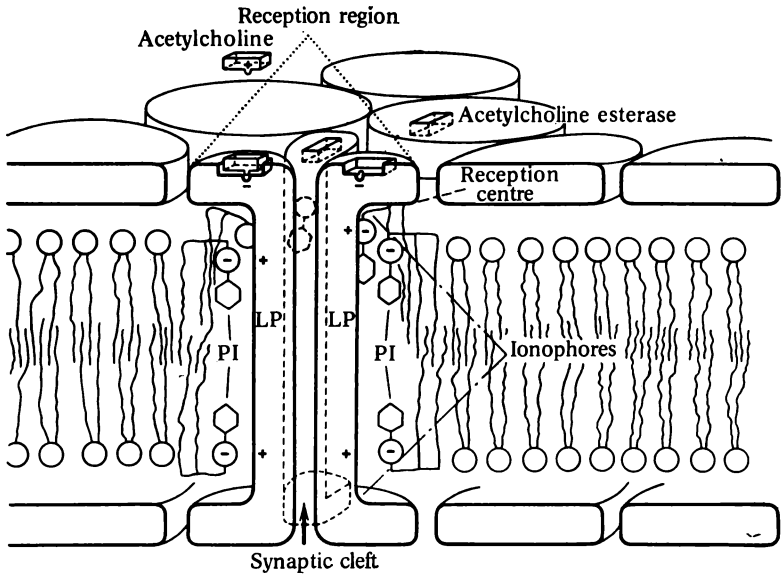


Fig. 11.21. A schematic representation of the macromolecular organization of the postsynaptic membrane according to De Robertis.
LP—lipoprotein; PI—phosphatidylinositol.

of quantum release. The postsynaptic membrane potential, which arises under the influence of a nerve impulse, is composed of a large number of miniature end-plate potentials created by individual portions—ACh quanta. The number of "quanta" participating in the reaction at a single myoneural junction depends on the presence of Ca^{2+} and Mg^{2+}: the Ca^{2+} ions stimulate the secretion of acetylcholine, and Mg^{2+} ions inhibit it.

It has been proved that each response to a nerve impulse is composed of a whole number of ACh "quanta" and the release of one "quantum" is an event with a very low probability ($p \ll 1$). The

number of nerve impulses, in which x portions ("quanta") of ACh are released, must be expressed by Poisson's law:

$$p_x = \frac{m^x}{x!} e^{-m} \quad .$$

where $x = 0, 1, 2, \ldots$ is the number of ACh portions; m is the average number of portions released by a single impulse. In the system studied

$$m = \frac{\text{Average response amplitude}}{\text{Average amplitude of spontaneous potentials}} = \frac{0.933 \text{ mV}}{0.4 \text{ mV}} = 2.33$$

The coincidence of the calculated and observed values of p_x has been found to be excellent. In particular, according to Poisson's law,

$$p_0 = \frac{n_0}{N} = e^{-m}$$

where n_0 is the number of null responses equal to 18 in the series considered; N is the total number of nerve impulses equal to 198;

$$m = \ln \frac{N}{n_0} = \ln \frac{198}{18} = 2.4$$

instead of the experimental value 2.33.

The "quantization" of the acetylcholine released compels us to believe that neurosecretion is a cooperative molecular process that may be treated on the basis of electronic-conformational interactions. The same ideas may prove useful for understanding of the nature of the generation of a nerve impulse in response to the action of a transmitter substance. Construction of appropriate kinetic models is a feasible task, but we have at our disposal insufficient information about the molecular structure of synaptic systems.

The problems of synaptic transmission are linked up with the investigation of neuron networks, by means of which the higher nervous activity is modelled. The existing models are mathematical rather than physical, since no sufficient knowledge of these phenomena is available at present.

Chapter 12 Mechanochemical Processes

12.1. Thermodynamics of Mechanochemical Processes

It has been shown in the preceding chapters how the chemical energy stored in the cell in ATP molecules is transformed to osmotic and electrical work. We shall now turn to mechanochemical processes, in which the chemical energy is converted to mechanical work.

Life is impossible without mechanical motion. The cells and organisms perform mechanical work, while moving as a whole and setting into motion their functional parts in a gravitational field by overcoming the resistance of an air or liquid medium, etc. Mechanical work is done under isothermal and isobaric conditions. It is thereby performed at the expense of chemical rather than thermal energy.

The conformational transformations of proteins imply the spatial movement of atoms making up macromolecules. If they are moving in the field of external forces, mechanical work is done. It can be done at the expense of the free energy liberated in the course of an enzymatic reaction. Hence, we may treat the conformational transformation of a protein as a mechanochemical process. However, as far as an enzymatic reaction in a solution is concerned, such an interpretation is meaningless. Any chemical reaction involves the motion of atoms, but this is not enough for such a reaction to be called a mechanochemical process.

Speaking of mechanical work, we imply a supermolecular macroscopic structure. If the molecules of an enzyme belong to the structure of such a system, then with its appropriate organization it can execute mechanical motion and perform work at the expense of the free energy of an enzymatic reaction. Hence, the working substances of mechanochemical systems in living organisms may be protein-enzymes. Moreover, they must be protein-enzymes. The source of mechanochemical work is chemical energy. But no biochemical reaction can proceed without enzymes. The mechanochemistry of living systems is enzymatic mechanochemistry.

A mechanochemical process can be accomplished by a cyclically operating machine which reverts to the initial state after each cycle.

The transitions of working substances from one chemical potential to another with the simultaneous performance of work occur in a medium that is external to the machine. The role of such a machine can be played by a polymeric polyelectrolyte fibre, whose length is changed with change of the pH of the medium.

The overall change of the internal energy of the system is

$$dE = T\,dS - dW + \sum \mu_i\,dn_i + \psi\,de + \ldots \qquad (12.1)$$

Here dW is the work done by the system; dn_i is the amount of substance i delivered from the reservoir into the system at a chemical potential μ_i; $\psi\,de$ is the electrical work. We shall confine ourselves to thermal, mechanical, and chemical processes, omitting $\psi\,de$ and the subsequent terms of the equation. We shall consider a homogeneous fibre of length l stretched by a force f. We have

$$dE = T\,dS - p\,dV + f\,dl + \sum_i \mu_i\,dn_i \qquad (12.2)$$

The integrated form of this equation is

$$E = TS - pV + fl + \sum_i \mu_i n_i + \mu_q n_q \qquad (12.3)$$

where n_q is the number of molecules that make up the fibre.

Suppose that there is only one reacting component. Hence, at constant p and T the system has two degrees of freedom, say, f and μ. The operating cycle of the mechanochemical machine may be depicted on the f-l plane. To each value of μ there will correspond a curve of l vs. f, which may be called an *isopotential* (like an isotherm on the p-V plane for a heat engine). All the points of such a curve can be obtained from a mechanical experiment carried out at a constant value of μ. Conversely, the l (f) curve at a constant value of n is similar to an adiabatic. Such a curve is called an *isophore*. On the μ-n plane the isopotentials and isophores are represented by straight lines parallel to the coordinate axes (Fig. 12.1). The curves of μ vs. n on such a plane may be called *isotonic* at constant force and *isometric* at constant length of the specimen.

The cycle on the f-l plane is shown in Fig. 12.2. The work is given by the following expression:

$$W = -\oint f\,dl = \int_1 \mu\,dn + \int_2 \mu\,dn = (\mu_1 - \mu_2)\,\Delta n \qquad (12.4)$$

This expression is similar to the expression for the work in the Carnot cycle:

$$W = \oint p\,dV = (T_1 - T_2)\,\Delta S$$

An essential difference is that the mechanochemical machine cannot be characterized by the efficiency of the cycle similar to

$$\eta_{\text{heat}} = (T_1 - T_2)/T_1$$

since the chemical potential has no absolute zero. The efficiency of the mechanochemical process may be represented by the ratio of the work performed in a real cycle to the work of an ideal reversible cycle:

$$\eta' = -\oint f\, dl / \oint \mu\, dn \tag{12.5}$$

Let us now consider nonequilibrium thermodynamics. The extension

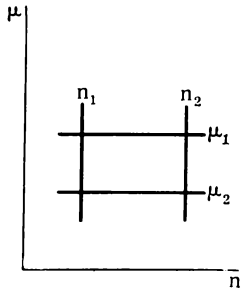


Fig. 12.1. The mechanochemical cycle on the μ, *n* plane.

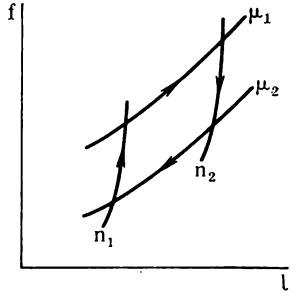


Fig. 12.2. The mechanochemical cycle on the *f*, *l* plane.

of a polymeric mechanochemical fibre in an isobaric and an isothermal process causes a change in free energy:

$$(dG)_{T,\,p} = f\, dl - \mathcal{A}\, d\xi \tag{12.6}$$

where $\mathcal{A}$ is the affinity; ξ is the chemical reaction coordinate (the extent of reaction or the degree of advancement of a chemical reaction). Equation (12.6) defines the internal mechanochemical force and the affinity:

$$f = \left(\frac{\partial G}{\partial l}\right)_{\xi,\,p,\,T} \tag{12.7}$$

$$\mathcal{A} = -\left(\frac{\partial G}{\partial \xi}\right)_{l,\,p,\,T} \tag{12.8}$$

In the equilibrium state these quantities are equal to zero. We have:

$$\left(\frac{\partial G}{\partial l}\right)_{\xi} = f = \left(\frac{\partial^2 G}{\partial l^2}\right)_{\xi}^{\text{eq}} \delta l + \left(\frac{\partial^2 G}{\partial l\, \partial \xi}\right)^{\text{eq}} \delta\xi = \left(\frac{\partial f}{\partial l}\right)_{\xi} \delta l + \left(\frac{\partial f}{\partial \xi}\right)_{l} \delta\xi$$

$$\left(\frac{\partial G}{\partial \xi}\right)_{l} = -\mathcal{A} = \left(\frac{\partial^2 G}{\partial \xi\, \partial l}\right)^{\text{eq}} \delta l + \left(\frac{\partial^2 G}{\partial \xi^2}\right)_{l}^{\text{eq}} \delta\xi = -\left(\frac{\partial \mathcal{A}}{\partial l}\right)_{\xi} \delta l - \left(\frac{\partial \mathcal{A}}{\partial \xi}\right)_{l} \delta\xi$$

Here the superscript "eq" stands for the equilibrium state; δl and $\delta\xi$ are the deviations of l and ξ from equilibrium values. The relations written above yield:

$$\left(\frac{\partial f}{\partial \xi}\right)_l = \left(\frac{\partial^2 G}{\partial \xi\, \partial l}\right)^{eq} = -\left(\frac{\partial \mathcal{A}}{\partial l}\right)_\xi \qquad (12.9)$$

It means that if $(\partial f/\partial \xi)_l$ is different from zero, the affinity $\mathcal{A}$ will depend on the length l.

Katchalsky and Oplatka have constructed a continuously working machine, in which a polyelectrolyte fibre (collagen) is alternately

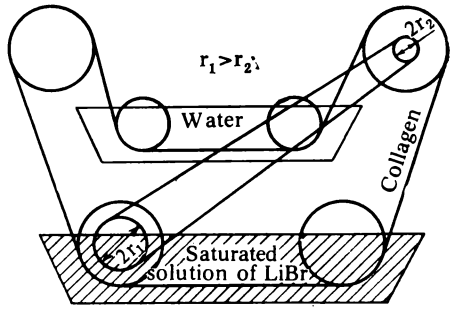


Fig. 12.3. The Katchalsky-Oplatka mechanochemical machine.

immersed in a salt (LiBr) solution and in pure water (Fig. 12.3). The operation of the machine is stopped when, as a result of the transfer of small ions by the fibre into water, the chemical potentials of both reservoirs are equilibrated. When the operation of the machine is reversed, it can serve for extracting the salt from the solution.

The polyanionic fibre is shortened with decreasing pH, say, upon addition of HCl. The increase of n_{HCl} lowers l at constant force f, i.e.,

$$\left(\frac{\partial l}{\partial n_{HCl}}\right)_f < 0$$

Conversely, under isometric conditions the addition of HCl increases the force:

$$\left(\frac{\partial f}{\partial n_{HCl}}\right)_l > 0$$

The reverse effect consists in changing the pH of the medium surrounding the polyanionic fibre if it is stretched. Thus,

$$-\left(\frac{\partial l}{\partial n_{HCl}}\right)_f = \left(\frac{\partial \mu}{\partial f}\right)_{n_{HCl}} > 0$$

The mechanochemical nature of a biological contractile system was disclosed by Engelhardt and Lyubimova (1939), who studied *myosin*, a protein of the muscle fibre (see page 413). When ATP is added, there occurs a reversible contraction of myosin fibres.

The mechanochemistry of polyelectrolytes is governed by the conformational transformations of macromolecules. These transformations are cooperative. The dissociation constants of the groups being ionized in a polymer differ from those in a monomer because of the electrostatic repulsion of the neighbouring charged groups which depends on the chain conformations. These phenomena are reflected in the curves of potentiometric titration (see Sec. 3.7).

At not too small forces f sufficient for the polyelectrolyte chain as a whole to be oriented, the chain length is expressed by the following equation (see Sec. 3.1):

$$\bar{x} = \frac{f}{3kT} \overline{h_0^2} \tag{12.10}$$

The mean square distance between the chain ends, $\overline{h_0^2}$, depends on the degree of ionization α, i.e., on the pH of the medium. If $\overline{dh_0^2}/d\alpha > 0$, the polyelectrolyte chains are stretched upon ionization under the action of a constant force and contracted with a decrease in the degree of ionization. If the sign of the derivative is negative, the effect of the pH of the medium is reversed. The second case occurs in synthetic polypeptides—the size of the coil is reduced with increasing α, since the increase of the degree of ionization destroys the helical structure. In the general case

$$\overline{h_0^2} \approx (\overline{h_0^2})_{\alpha=0} \exp\left(-\frac{\alpha^2 \Delta E}{kT}\right) \tag{12.11}$$

where ΔE is the difference in the energy of interaction of free charges between the coiled and extended chain conformations; ΔE is the measure of cooperativity of the system: if ΔE is equal to zero, the chain is noncooperative and $\overline{h_0^2}$ is independent of the degree of ionization α. The sign of the derivative $\overline{dh_0^2}/d\alpha$ depends on the sign of ΔE. The same theory permits one to find out how the pH of the medium is changed on extension of the chain.

The above-described cooperative mechanism is not the only one; polyelectrolyte mechanochemical processes determined by changes in the degree of ion binding are also possible.

Thus, the changes in the chain conformations brought about by the changes of the chemical environment under the action of a constant force perform mechanical work. In turn, the acting force must alter the chain conformations. Theory shows that if the applied force is not too large, it stabilizes the helical conformation of the chain. Conversely, a large force stabilizes the conformation of a strongly

extended coil. In charged chains, not too large an external force causes an increase in the degree of dissociation which brings about helix-coil transitions at a fixed temperature. With a large force the degree of dissociation that causes such a transition may be lower than in the absence of a force. Depending on the magnitude of the applied force, the helix-coil transition may involve an increase or a decrease in the chain size in the direction of the force.

Biological contractile systems, which accomplish mechanochemical processes, are far from being simple polyelectrolyte models. However, the properties of polyelectrolytes are essential for the understanding of mechanochemical phenomena.

There is every reason to believe that the chemical energy required in mechanochemical processes is supplied by macroergic substances, adenosine triphosphate (ATP) in the first place. The hydrolysis of ATP occurs with the participation of ATPase. The working substances for mechanochemical processes are contractile proteins. The discovery of the ATPase activity of one of such substances, the myosin of the muscle, made by Engelhardt and Lyubimova, is the pivotal point in the entire biological mechanochemistry.

In dealing with biological mechanochemical processes we shall focus our attention mainly on muscle contraction.

12.2. The Structure of the Muscle and Muscle Proteins

All vertebrates, including man, have three distinctly different types of muscle—smooth muscles in the walls of hollow organs; the cross-striated cardiac muscles, and the cross-striated skeletal muscles. We shall mainly be concerned with skeletal muscles.

Muscles have a fibrillar structure. Under the ordinary microscope we can, without difficulty, observe the striated structure of muscle fibres. An individual muscle fibre has a diameter of 0.02-0.08 mm (20-80 μm). It is surrounded by a membrane about 0.01 μm thick. The fibre is formed by 1000-2000 thin strands of contractile material, *myofibrils* 1-2 μm in diameter. The fibrils have a sheath formed by tubules and vesicles of the *sarcoplasmic reticulum*. The microscopic structure of the myofibril is shown in Fig. 12.4. Myofibrils are in turn made of bundles of protein strands—thick and thin filaments. The symmetry of their arrangement in the cross-section is hexagonal (Fig. 12.5). Figure 12.6 shows the longitudinal cross-section of the myofibril; its structure is schematically depicted in Fig. 12.7. The thick lines in Fig. 12.6 (they are also clearly seen in Fig. 12.4) represent the so-called *Z-lines*—the longitudinal cross-sections of Z-discs. A package of myofibrils between two Z-lines is called a *sarcomere*. The sarcomere is divided into several zones, which are clearly observed under the polarizing microscope. The central line A is anisotropic; it exhibits birefringence. It is flanked by two isotropic

I-bands. On stretching of the resting muscle, in the middle of the A-band there appears an H-zone of lesser density. These structural details are shown schematically in Figs. 12.6 and 12.7.

Electron-microscopic studies carried out by Huxley and Hanson disclosed the arrangement of thick and thin protein strands (myofilaments) in the sarcomere (Figs. 12.5 and 12.7). The thick filaments are made of the protein myosin; the thin filaments are composed

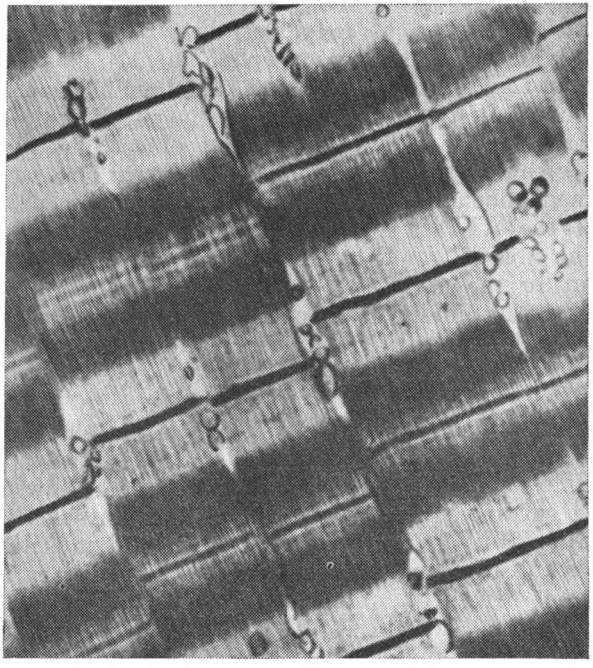


Fig. 12.4. An electron micrograph of the myofibril of the frog muscle.

of a second protein, *actin*. Each thick filament consists of 180-360 longitudinally oriented myosin molecules responsible for the anisotropy of the dense A-band. The less dense I-band is formed by thin filaments of the protein actin, whose molecules are double helices (the *F-form of actin*) formed by the polymerization of the globular *G-actin*. In the sarcomere the number of G-globules is about 800 per one thin filament. The thin filaments of F-actin pass through the Z-discs.

The actin filament is a double helix, with G-subunits recurring every 5.46 nm along each of the two filaments with a distance of

36-37 nm between the points of intersection. The diameter of the double helix of actin is 6-8 nm. The thick filaments of myosin are 14 nm in diameter. Protruding from the myosin filament are the

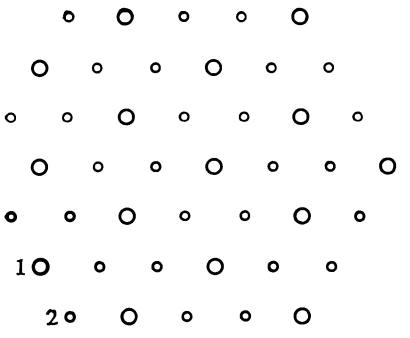


Fig. 12.5. Schematic representation of a portion of a cross-section of the myofibril:
1—thick filaments; *2*—thin filaments.

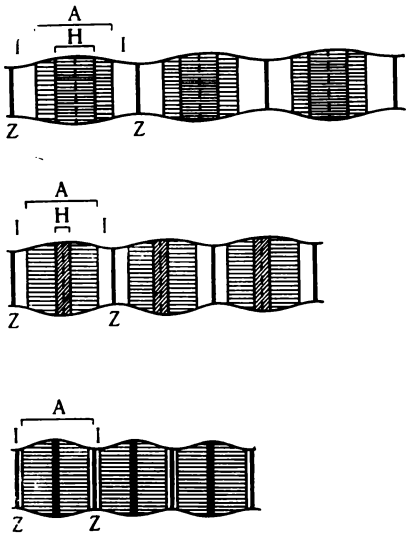


Fig. 12.6. The longitudinal cross-section of the myofibril with three different lengths.

"heads" which are arranged in a helical manner on the filament. At a given level, two protruding heads are opposite each other. The next two protruding heads are a distance of 14.3 nm apart and are turned relative to the first pair at an angle of 122°. The structure as a whole recurs with a period of $3 \times 14.3 = 42.9$ nm.

Myosin is a fibrillar protein—it consists of molecules 150 nm in length and about 2 nm in diameter. The thickened end of the molecule ("head") has a length of about 20 nm and a width of about 4 nm. A detailed electron-microscopic study of myosin has led to the results presented in Fig. 12.8. The myosin molecule is composed of *light meromyosin* (LMM) and *heavy meromyosin* (HMM), the latter being made up of two fragments, S-1 and S-2. The molecular masses and sizes of these fragments are given in Fig. 12.8. The long

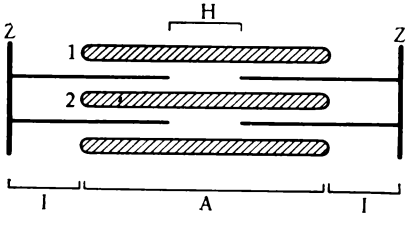


Fig. 12.7. The structure of the sarcomere (schematic).

right part of the molecule is a double superhelix. The total molecular mass of myosin is 5×10^5, and the degree of α-helicity is high (about 58 percent).

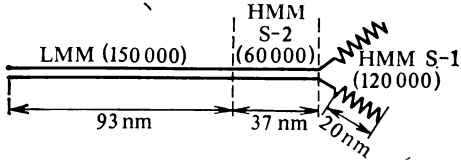


Fig. 12.8. The structure of the myosin molecule (schematic). The longitudinal dimensions of fragments and their molecular weights are indicated.

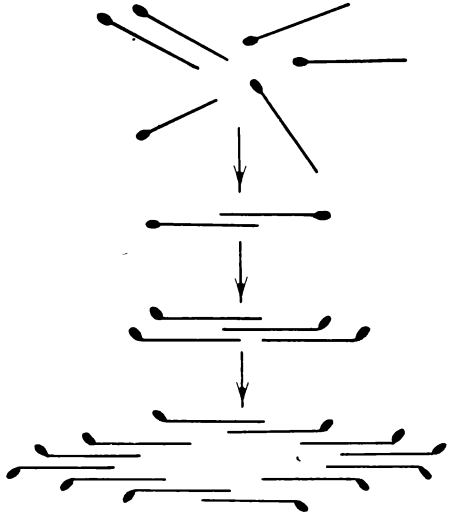


Fig. 12.9. Schematic representation of the aggregation of myosin molecules.

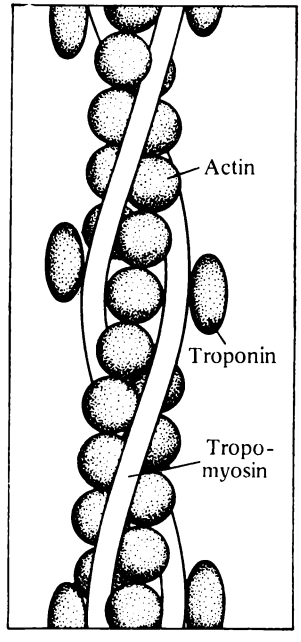


Fig. 12.10. The structure of the thin filament (schematic).

During the formation of thick filaments the myosin molecules undergo aggregation, presumably as a result of electrostatic interactions between their tails (LMM). The pattern of aggregation is shown in Fig. 12.9. The heads protruding from a thick filament are made up of the S-1 fragments of HMM.

The spheroidal monomers of G-actin have a diameter of about 5.5 nm and a molecular mass ranging from 46,000 to 47,000. The degree of α-helicity is about 30 percent. The polymerization of G-actin to F-actin involves ATP and takes place as follows:

$$n(\text{G-actin-ATP}) \rightleftharpoons (\text{F-actin-ADP})_n + nP_i$$

The polymerization in solution proceeds in the presence of salts, in particular Mg^{2+}.

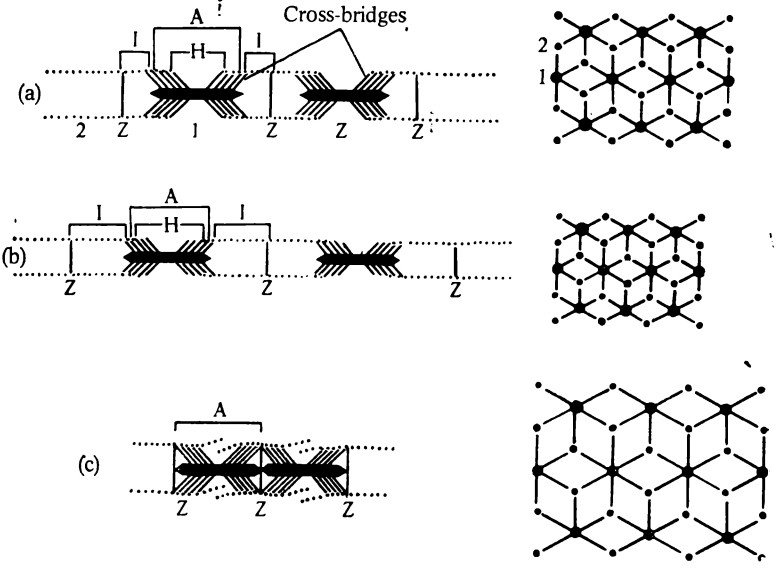


Fig. 12.11. The sliding filament mechanism of muscle contraction:
(*a*) The normal structure of the fibre. (*b*) The extended fibre. (*c*). The strongly shortened fibre. Right: a cross-section.

In thin filaments, apart from actin, there are present other proteins—*tropomyosin* and *troponin*. Tropomyosin has a molecular mass of about 70,000; it is composed of molecules about 45 nm in length with the long-to-short axis ratio of an effective ellipsoid being larger than 20. The degree of α-helicity of tropomyosin is about 90 percent. Troponin is a globular protein made up of three components with molecular masses 37,000, 23,000, and 19,000 (the sum total being equal to 79,000). The molecules of tropomyosin are arranged along each of the grooves of the double helix of actin, the troponin globule residing near the end of each tropomyosin molecule. One molecule of tropomyosin is in contact with seven G-actin molecules (Fig. 12.10).

So far we know little about other regulatory proteins of myofibrils—about α- and β-actinins and M-protein.

Vazina and coworkers have established that various contractile proteins form in solution anisotropic liquid-crystalline structure capable of polymorphic transformations. As we have seen (page 351), these properties determine the functionality of biological membranes. One might expect that these liquid-crystalline properties of contractile proteins are essential to muscular contraction.

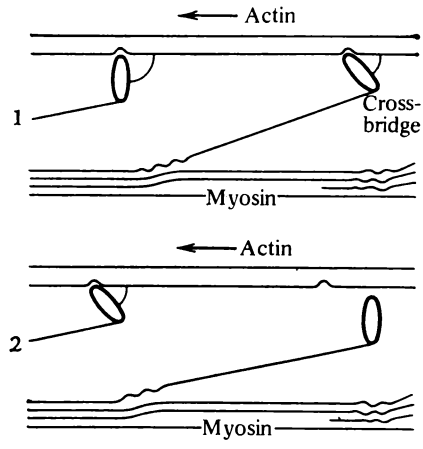


Fig. 12.12. The change in the angle between the cross-bridges and the thin filament.

1 and *2*—successive moments of motion; A—actin; M—myosin.

Light and electron microscopy have enabled the elucidation of the basic structural features of muscle contraction.

When the muscle contracts (is shortened), the I-bands are narrowed without change of the extent of the A-band; the Z-discs move to each other. Finally, the I-bands disappear entirely, and in the centre of the sarcomere there appears a compacted portion. The volume of the sarcomere is changed but slightly on contraction; hence, the sarcomere becomes thicker. This is shown schematically in Fig. 12.6.

Electron micrographs show that during contraction the thick and thin filaments telescope together by sliding past each other, as a result of which the sarcomere becomes shortened. This is the *sliding filament model* of muscle contraction developed by H. Huxley.

The interaction of thick and thin filaments occurs via the HMM heads of myosin which form *cross-bridges* that link these filaments. In the normal physiological state of the muscle, the thick and thin filaments overlap in such a manner that all possible bridges can be formed (Fig. 12.11*a*); this does not occur when the muscle contracts strongly (Fig. 12.11*b*). With large shortening the thin filaments undergo deformation (Fig. 12.11*c*).

Myosin is responsible for the anisotropy of the A-band: in the I-bands there is no myosin present and, hence, the thin filaments themselves do not produce anisotropy.

In the sliding filament model each bridge operates cyclically. The bridge pushes or pulls actin to the centre of the A-band to a distance of the order of 5-10 nm; then it splits off from actin and

again attaches itself to actin at another point which was initially at a larger distance from the centre of the A-band. The cycle then repeats. The continuous movement of the actin filaments occurs as a result of the asynchronous action of the bridges.

The pushing or pulling force developed by the bridge may arise only as a result of a conformational change. Such a change may bring about either an active alteration of the angle at which a HMM head is attached to a thin filament or a change of its shape. Experiment shows that the connection between bridges and thick filaments is flexible and their linkage with thin filaments is very rigid. Figure 12.12 presents the model of displacement of a thin filament caused by the change of the angle at which a bridge is attached to it.

When the muscle is in the rigor state, there are formed rigid and immobile links between bridges and thin filaments. The helical periodicity is upset, the 42.9 nm spacing disappears entirely and is replaced by layer lines which can be indexed either on a 36 nm or 38 nm helical repeat. On the other hand, the 14.3 nm spacing remains unchanged. These phenomena can also be accounted for by changes in the disposition of the bridges with the main backbone of the thick filament remaining intact.

The above-described structural picture, which is consistent with the sliding model, has been convincingly verified. The physical theory of muscle contraction must be based on this model taken as an experimental fact. The molecular interpretation of muscle contraction must take into account the results of biochemical and physical investigations in addition to structural findings.

12.3. The Chemistry and Physics of the Muscle

The biochemistry of the muscle is based on the enzymic activity of myosin (see page 410)—its ability to catalyze the hydrolytic cleavage of ATP. The scheme of the reaction is as follows:

$$M + ATP \rightleftharpoons M\text{-}ATP \rightleftharpoons M\begin{matrix}\nearrow ADP \\ \searrow P\end{matrix} \rightleftharpoons M + ADP + P_i$$

The reaction strongly depends on the ionic medium, on the presence of divalent cations. In the muscle it occurs at pH = 7.4. For the reaction to proceed *in vitro* it is necessary that Ca^{2+} ions be present; it is optimal with the concentration of these ions being equal to the concentration of ATP or somewhat greater. The Mg^{2+} ions inhibit the ATPase activity of myosin *in vitro*.

In the muscle, myosin functions upon interaction with actin—as has been pointed out, through HMM bridges—in the actomyosin complex. Under these conditions the Mg^{2+} ions exert an activating

effect. Thus, for the ATPase activity to be exhibited *in vivo*, both Ca^{2+} and Mg^{2+} ions are required.

The activation of the striated muscle occurs as a result of the transmission of a nerve impulse to it. Muscle contraction can also be initiated by an artificial electrical pulse (Galvani experiments). The action of the impulse leads to an increase in the concentration of Ca^{2+} ions which interact with a fibril. Each fibril is enwrapped by a complex system of longitudinal and transverse thin tubules and vesicles known as the *sarcoplasmic reticulum*. It has been proved, with the aid of labeled ^{45}Ca atoms, that the transmission of the nerve impulse causes the extrusion of Ca^{2+} ions from the reticulum into the gel-like cytoplasm (the *sarcoplasm*) which surrounds the protein filaments. The concentration of free Ca^{2+} ions in a relaxed muscle is very low, probably less than 10^{-7}—10^{-8} *M*. Upon activation it increases by 2 or 3 orders.

The reaction given on page 417 is accompanied by the subsidiary reaction called the *Lohmann reaction*:

$$\text{ADP} + \text{Phosphocreatine} \underset{\longleftarrow}{\overset{\text{Creatine kinase}}{\longrightarrow}} \text{ATP} + \text{Creatine}$$

It has been proved that ATP is split just upon contraction of the muscle. Fluorodinitrobenzene inhibits creatine kinase and stops the Lohmann reaction. At the same time, it does not affect muscle contraction and the first reaction. It has been found, with the aid of this enzyme, that ATP is split during a single contraction (a twitch) and tetanus contraction. On the other hand, the reverse reaction of formation of phosphocreatine from creatine can be blocked by iodoacetate. In such a case, the consumption of phosphocreatine characterizes the amount of ADP formed during contraction. The amounts of ATP split on contraction correspond to the loss of phosphocreatine in the Lohmann reaction.

The relaxation of the muscle to its initial state is determined by the expulsion of Ca^{2+} ions from the sarcoplasm into the reticulum.

Thus, the Ca^{2+} ions regulate the contraction processes in muscles. Presumably, this occurs with the direct participation of troponin and tropomyosin (see page 415). In the absence of Ca^{2+} ions troponin complexed with tropomyosin inhibits the interaction of actin with myosin bridges. The calcium that enters the sarcoplasm binds to troponin and suppresses its inhibiting action. Hence, Ca^{2+} ions play the role of a derepressor and switches the thin filament over from the inactive to the active state.

Let us now turn to the mechanical properties of the muscle which have been studied in a series of biophysical investigations carried out for many years by Hill and other authors.

The term "contraction" applies to the muscle in two senses. *Isometric contraction* occurs in the muscle when it has a fixed length. In

this case, a tension develops without shortening. Conversely, *isotonic contraction* means that the muscle is allowed to shorten against a constant load. In this case, the muscle performs work, lifting the load.

The experimental procedure of corresponding measurements was developed in the classical works of Hill and later repeatedly improved. Figure 12.13 shows the experimental setup. The muscle *M* is attached to one end of the lever. The other end of the muscle is

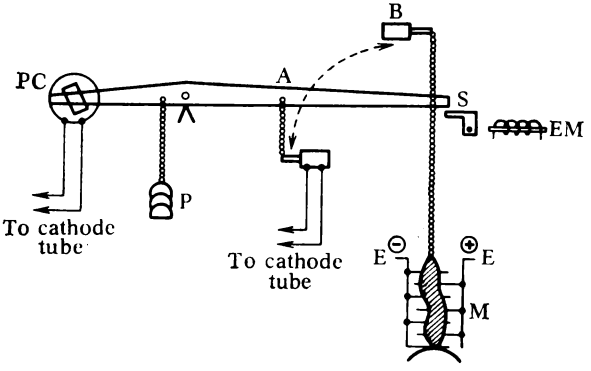


Fig. 12.13. Schematic of the setup for studying muscle contraction.

tied to a rigid support. Contraction is stimulated by electrodes *E*. When isotonic twitches are to be studied, the muscle is loaded at position *P*. The tension is recorded by means of a tension transducer *T* at position *A*. When isometric tension is measured, the transducer is switched over to position *B*. An isotonic shortening is determined by the movement of the other end of the lever recorded by means of a photocell PC. The stop *S* controlled by an electromagnet *EM* is used to hold up a contraction at any desired length or to release an isometrically contracting muscle to any desired extent. Experiments with the frog sartorius muscle show that during the first 15 ms after the excitation of an isometric contraction there occur a series of changes determined by the release of Ca^{2+} ions from the sarcoplasmic reticulum. Here the tension begins to increase, reaching its maximum after 170 ms (at 0°C). Then it declines, disappearing completely after more than 1 s. During an isotonic contraction the shortening is diminished with increasing load *P*; its maximum is attained the earlier the larger the load. Then relaxation to the initial state takes place. The development of an isotonic tension in the muscle practically follows the same time course as the development of an isometric tension.

Hill has empirically established the basic, characteristic equation in the mechanics of muscle contraction. It relates the steady velocity of an isotonic contraction (the velocity of shortening) V to the force (load) P and has a hyperbolic form:

$$(P + a)\, V = b\, (P_0 + P) \qquad (12.12)$$

where P_0 is the maximum isometric tension developed by the muscle or the maximum load against which the muscle supports without shortening; a and b are constants. At $P = 0$ the velocity of shortening is maximal:

$$V_{max} = b \frac{P_0}{a} \qquad (12.13)$$

Conversely, at $P = P_0$ we have $V = 0$. At $P > P_0$ the muscle is no longer shortened; it is stretched. This stretching does not obey Eq. (12.12).

The constant b, which has the dimensions of a velocity, increases rapidly with temperature: on heating by 10°C near 0°C the constant b is doubled. The constant a varies from 0.25 P_0 to 0.4 P_0. Hence, V_{max} is 2.5-4 times greater than b.

Hill's equation is valid only at a constant velocity of shortening. At the same time, the scope of its application is limited by muscle lengths close to physiological lengths. The maximal tension P_0 depends on the length of the sarcomere. Figure 12.14 gives the dependence of P_0 on the relative sarcomere length in the muscle fibre: P_0 is maximal and corresponds to a certain plateau region in Fig. 12.14 at lengths close to physiological lengths. In this case, there is observed a complete overlap of actin filaments and the HMM heads of myosin filaments and, hence, there is a possibility of formation of a maximum number of bridges (see Fig. 12.11*a*). With large sarcomere lengths the overlap and the number of possible bridges declines (Fig. 12.11*b*) and P_0 also decreases. Finally, with extensive shortening the filaments are deformed (Fig. 12.11*c*) and P_0 again drops. Hill's equation is valid in the plateau region, in the region of largest P_0 values, i.e., with sarcomere lengths ranging approximately from 1.7 to 2.5 μm or at relative lengths from 0.9 to 1.25.

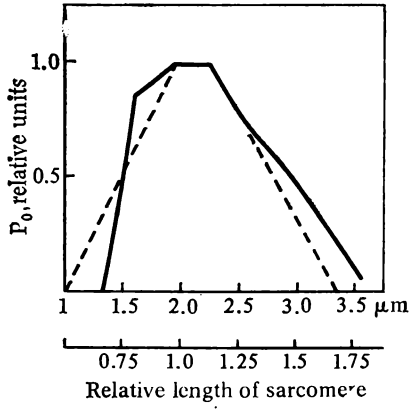


Fig. 12.14. The dependence of the maximum force on the length of the sarcomere. The solid line represents the data of Edman, and the dotted line, the data of Shenberg and Podolsky.

The quantity P_0 depends little on temperature, increasing only slightly with rise of temperature. This also refers to the constant a, which is proportional to P_0.

The mechanical properties of the muscle are studied during a twitch or tetanic contraction. A single impulse causes a single twitch. With sufficiently frequent impulses delivered one after another, say, with 15 impulses per second, the single twitches fuse to give a tetanus, since each of the successive impulses falls within the refractory period of the previous one (see page 381). The work done by the muscle on contraction is equal, according to Hill's equation, to

$$W = PVt = bPt\,\frac{P_0 - P}{P + a} \tag{12.14}$$

The function $W(P)$ is bell-shaped; W turns to zero at $P = P_0$ and $P = 0$ and reaches a maximum at $P = [a\,(P_0 + a)]^{1/2} - a$ and since $a \approx 0.25P_0$ at $P \approx 0.31P_0$.

According to the sliding-filament theory, the tension developed by the muscle is entirely determined by the filaments of actin and myosin and Z-discs. All these elements are not quite rigid, being to some extent compliant. The terminal sarcomeres of the muscle fibre are attached to the connective tissue of the tendons and here there is also observed compliance, plasticity. Simultaneously, these elements impart some elasticity to muscle movement. The total contribution of elastic and plastic deformations does not, however, exceed 3 percent of the tension developed by the muscle. The muscle should, however, be treated as a viscoelastic body. As we shall see later, Hill's equation describes only viscous flow in the muscle.

The study of the viscoelastic properties of the muscle is important for the elucidation of its kinetic behaviour under nonsteady-state conditions (see Sec. 12.5).

The maximal tension P_0 of the frog sartorius muscle is approximately equal to 30 N/cm^2, which corresponds to 3×10^{-10} N for one thin filament. Considering that each single element contains one myosin molecule and each closing-and-opening cycle of the bridge is accompanied by the splitting of one molecule of ATP, whose energy is used to the extent of 50 percent, we obtain the following characteristics of the single component: force, 3×10^{-12} N; distance, 10^{-6} cm; energy, 6×10^{-20} J (15 kT); time, 1 ms.

While the contracting muscle is doing work, it liberates heat. The heat is also released on isometric contraction. During relaxation of the isotonic tension the falling load performs work on the muscle which is also converted to heat.

At the early stage of the contraction process, before the tension has developed or before shortening has occurred, an activation heat Q_a of the order of 4.2 mJ per 1 g of muscle mass is liberated. This

heat is presumably associated with the entry of Ca^{2+} ions into the sarcoplasm and with their interaction with the actomyosin system. Then, as the muscle is contracting and work is being done the heat of contraction Q_c is liberated. If the muscle is shortened, the heat Q_c is liberated faster than on isometric contraction during the same time. The total change of energy during the contraction process is given by

$$\Delta E = Q_a + Q_c + W \tag{12.15}$$

The rate of heat liberation is maximal in the region of the physiological dimensions of the muscle. Probably, Q_c is associated with the maintenance of tension and arises as a result of the interaction between thick and thin filaments.

In the Hill experiments, the muscle was continuously stimulated under isometric conditions, after which the muscle was suddenly released to a pre-determined shorter length, lifting the load. When this is done, it is always found that extra heat is given out over and above the isometric amount—the heat of shortening. This extra heat is proportional to shortening at constant load and to the load at constant amount of shortening. Under steady isotonic conditions the power of the muscle, i.e., the rate of energy liberation in the form of heat and work, is equal to

$$\dot{E} = \dot{W} + \alpha V + m = (\alpha + P)\, V + m \tag{12.16}$$

where m is the rate of heat liberation on isometric tetanus contraction; αV is the rate of liberation of extra heat.

It has been found that if the muscle is capable of shortening, the total energy liberated by the muscle during a single twitch is greater than on isometric contraction. This is the *Fenn effect.*

The mechanical efficiency of the muscle can be defined as the ratio of work done to the energy used:

$$\eta' = \frac{W}{E} = \frac{W}{W+Q} \tag{12.17}$$

In the contractile phase η' reaches 45 percent in a frog and 75 percent in a tortoise. In an ideal cyclical mechanochemical process that does work, the free energy remains constant:

$$\oint \Delta G = 0$$

or, according to Eq. (12.3),

$$\oint f\, dl = -\oint \mu\, dn \tag{12.18}$$

[cf. Eq. (12.5)]. The value of η' characterizes the deviation from Eq. (12.18), i.e., the degradation of the chemical energy into heat.

12.4. The Theory of Muscular Contraction

The molecular theory of muscle contraction has not yet been worked out. A number of facts, however, have been satisfactorily interpreted.

Oplatka advanced the theory of steady muscle contraction within the framework of nonequilibrium thermodynamics (1972). Muscle contraction is treated as a plastic flow with friction. Hill's equation (12.12) is used as an experimental fact and a relation is established between the speed of shortening V and the rate of ATP splitting.

A number of works are concerned with the electrostatic effects and make use of the polyelectrolyte model of muscle proteins. It is difficult to reconcile the possible electrostatic effects in a medium, which is a decinormal salt solution, with the observed large values of P_0.

Szent-Györgyi proposed in his time a contraction hypothesis based on the quantum migration of energy along the quasi-crystalline network of water that surrounds myosin. These ideas are entirely speculative. Much later, McClare again addressed to the consideration of the resonance transfer of "molecular energy" in the muscle (1972). Davydov presumed that energy transfer occurs by way of the propagation of a *soliton*—a nondissipating wave of vibrational excitation of the C=O bonds in the α-helices of myosin. As a result, the thick filament is bent, causing the closure of the bridges (1973). These hypotheses have not yet been experimentally verified.

Tonomura proposed a spectacular molecular interpretation of the sliding filament model (see page 416). The scheme suggested by Tonomura is given in Fig. 12.15. It is presumed that: (1) muscle contraction is associated with the phosphorylation and dephosphorylation of myosin; (2) the conformation of the head of myosin is changed upon addition of ATP; (3) the F-actin-myosin linkage is split with the formation of a myosin-ATP complex at high ATP concentrations and of a myosin-phosphate-ADP complex at low ATP concentrations; the strength of the linkage depends on the conformation of myosin and of a complex of F-actin with a regulatory protein, whose conformation is controlled by Ca^{2+} ions; (4) the complete cycle occurs as a result of the splitting of the myosin-F-actin linkage after the conformational change of the myosin head caused by the liberation of energy. The linkage is re-formed when the myosin molecule regains its initial form.

The conformational changes in myosin and actin under the influence of ATP and in the contracting muscle have been established by many methods. They have, however, been studied not sufficiently for the model similar to the one described above to be substantiated.

We have already mentioned the liquid-crystalline properties of contractile proteins (see page 416). When the muscle shortens, the repeat period of the lattice made up of protofibrils is changed. When the lattice of thin filaments is inserted into the lattice of thick filaments, the tetragonal symmetry is replaced by hexagonal symmetry. This may be looked upon as a polymorphic transition in the liquid-crystalline system. The question requires further investigations.

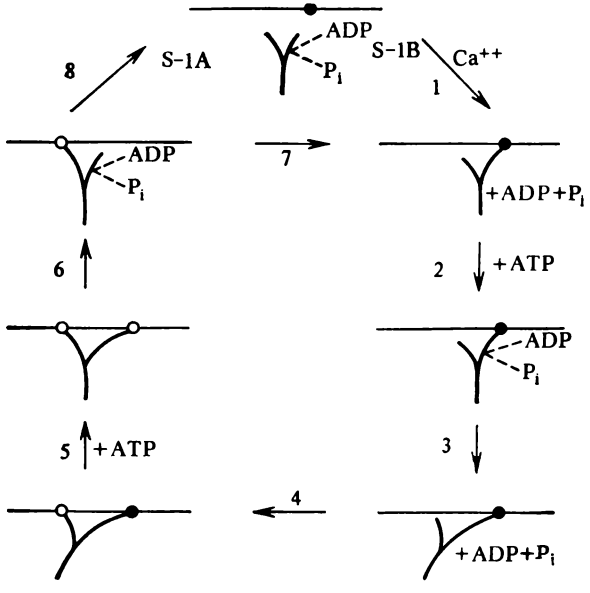


Fig. 12.15. Diagram showing the operation of the cross-bridge:
S-1B is the part of the "head" of myosin that contains an active site of ATPase and a binding site to F-actin; S-1A contains only a binding site to F-actin; *1*—arrival of Ca^{2+}, binding of S-1B; *2*—phosphorylation; *3*—dephosphorylation, a conformational transition, sliding; *4*—binding of S-1A; *5*—formation of a complex of myosin with ADP and P_i; *6*—dissociation of the bond of S-1B, a spontaneous conformational transition; *7*—the bond exchange S-1A → S-1B; *8*—rupture of the bridge.

The first attempt to work out a quantitative physical theory based on the sliding model was undertaken by A. Huxley (1957). He assumed that the active projection of HMM (designated as M, myosin) oscillates along the thin filament, while the active centre of actin (A) is immobile. The reactions involved are as follows:

$$A + M \rightarrow AM \text{ (formation of a bridge)}$$
$$AM + ATP \rightarrow A \cdot ATP + M \text{ (breaking of the bridge)}$$
$$A \cdot ATP \rightarrow A \cdot ADP + P_i \text{ (splitting of ATP)}$$

A kinetic problem is solved, the rate constants of the first two reactions being assigned certain dependences on the distance between A and M along the myofibril. The solution takes account of the relative movement of A and M during the shortening of the sarcomere. The numerical choice of a number of parameters leads to agreement with the results of calculation of P (V) by means of Hill's equation.

The Huxley theory was refined by Deshcherevsky (1968). Three states of the bridges are considered: closed bridges which develop a pulling force; closed bridges which retard the sliding of filaments; and open bridges. The bridges are closed, independently of each other, and pull the filaments, causing active contraction. The same bridges then retard the movement, since the filaments have been displaced and, finally, the bridges are broken. Let n_0 be the total number of active HMM heads in half of the thick filament, n the number of pulling bridges, and m the number of retarding bridges; and so we obtain

$$\dot{n} = k_1 (n_0 - n - m) - \frac{v}{L} n \tag{12.19}$$

$$\dot{m} = \frac{v}{L} n - k_2 m \tag{12.20}$$

Here v/L is the rate constant of the conversion of pulling bridges to retarding bridges which is equal to the velocity of relative movement of filaments, v, divided by the distance L between two neighbouring active centres of actin, on which the bridge can be closed. These equations are supplemented by the expression of Newton's second law:

$$M\dot{v} = f_0 (n - m) - f n_0 \tag{12.21}$$

where M is the mass moved; f_0 is the active force developed by a single pulling bridge; f is the external force (load) per one bridge. Equation (12.21) describes movement with friction expressed as $f_0 m$. Under steady-state conditions $\dot{n} = \dot{m} = 0$ and $\dot{v} = 0$. Omitting n, m, and n_0 from Eqs. (12.19), (12.20), and (12 21), we find that

$$(f + a)\, v = b\, (f_0 - f) \tag{12.22}$$

This equation coincides with Hill's equation (12.12). All the quantities in Eq. (12.22) are referred to a single bridge; v is the velocity of shortening in half the sacromere. The constants a and b are expressed through k_1, k_2, f_0, and L:

$$a = k_1 f_0 / (k_1 + k_2) \tag{12.23}$$

$$b = k_1 k_2 L / (k_1 + k_2) \tag{12.24}$$

Thus, we see that the Hill empirical equation can be derived theoretically. The equation corresponds to the steady sliding of filaments with the frictional force being proportional to the velocity.

We see that the Deshcherevsky cycle does not include the reverse processes. If these processes are taken into account, we obtain (see page 328):

$$k_1 k_2 \frac{v}{L} = k_{-1} k_{-2} k_{-3}$$

and if k_{-1}, $k_{-2} \ll k_1$, k_2, then $k_{-3} \gg v/L$. There must thus be a second cycle that drives the cycle under consideration. This second cycle is a closed chain of chemical transformations during which ATP is split.

This theory provides a correct solution for steady isotonic contraction but it does not account for the tension arising during isometric contraction.

We shall now expound the theory of steady contraction based on the theory of absolute reaction rates (Volkenshtein, 1969).

This theory rests on the sliding filament model with friction. A pulling or pushing force arises as a result of the closure of a bridge and of the succeeding events involving the conversion of the chemical energy of ATP to the energy of the conformational change of the protein contractile system. Each bridge develops tension f_0 when being closed. It is the closing and opening of the bridge that produces friction between filaments. Newton's second law will be rewritten in the form

$$M\dot{V} = P' - P - BV \tag{12.25}$$

where M is the mass being moved; P is the applied tension; P' is the tension developed; BV is the force of friction. Under steady-state conditions $\dot{V} = 0$. Let us represent P in terms of $n_0 f$, where n_0 is the maximum number of operating bridges. The bridges operate asynchronously, and at a given load only a certain effective number of bridges participate in the shortening of the muscle, $n = wn_0$, where $w < 1$. While being closed, these bridges develop tension $P' = nf_0 = wn_0 f_0$. So they are responsible for the internal friction and, hence, $B = \beta n = \beta w n_0$. The value of w depends on P, i.e., on f. We obtain for steady-state conditions:

$$wf_0 - f - \beta w v = 0 \tag{12.26}$$

whence

$$v = \frac{1}{\beta}\left(f_0 - \frac{f}{w(f)}\right) \tag{12.27}$$

The physical assumption consists in that $w = n/n_0$ is a function of the force f alone but not of the velocity v. The dependence of w on v results only from the fact that the force depends on the velocity. This supposition is independent of the law of forces (12.25) and (12.26). Let us find the shape of the function $w(f)$. At maximum

load P_0 the load is held by all the bridges, i.e. $n(f_0) = n_0 w(f_0) = n_0$ and $w(f_0) = 1$. Correspondingly, according to Eq. (12.27), at $f = f_0$ the velocity $v = 0$. At $f = 0$ the minimum number of bridges are working, $n(0) = n_0 w(0) = n_{\min}$ and $w(0) = n_{\min}/n_0 \equiv r < 1$. The maximum velocity of shortening at $f = 0$ is equal to

$$v_{\min} = f_0/\beta \tag{12.28}$$

We shall naturally assume that the tension developed by the muscle during an isometric contraction or the load which the muscle bears during an isotonic contraction is linearly dependent on the number of acting bridges, since each of them develops a constant tension f_0. Hence,

$$P = n_0 f = (A + Cn) f_0$$

At $f = f_0$ we have $n = n_0$ and $n_0 = A + Cn_0$; at $f = 0$ we obtain $n = n_{\min} = rn_0$ and $A = -Crn_0$. We find that

$$w = n/n_0 = r + (1 - r) f/f_0 \tag{12.29}$$

Substituting this expression into Eq. (12.27), we obtain Hill's equation in the following form:

$$v\left(\frac{r}{1-r} f_0 + f\right) = \frac{f_0}{\beta} \frac{r}{1-r} (f_0 - f) \tag{12.30}$$

Thus, the constants a and b are given by

$$a = \frac{r}{1-r} f_0 \tag{12.31}$$

$$b = \frac{r}{1-r} \frac{f_0}{\beta} = \frac{a}{\beta} \tag{12.32}$$

Experiment gives $a \approx 0.25\ f_0$. Hence, $r/(1 - r) \approx 0.25$ and $r \approx \approx 0.2$. From Eq. (12.30) it follows that

$$v_{max} = \frac{f_0}{\beta} = \frac{1-r}{r} b \approx 4b \tag{12.33}$$

The friction is determined by all the events that occur when the bridge is closed. The elementary act is now considered to involve the closing of the bridge, the hydrolysis of one molecule of ATP, the conformational change of the bridge, the performance of work $f_0 L$ (where L is the path of a single shortening) and the opening of the bridge. The reverse microscopic act leads to the synthesis of one molecule of ATP from ADP and P_i and to the performance of work $f_0 L$ on the system.

We shall make use of the theory of internal friction of liquids developed by Eyring (page 189). A change in the state of the system may be represented by the free-energy curve shown in Fig. 12.16. Here *1* is the state of the system before the bridge is closed; *2* is the

state of the system after the bridge is closed and after all the succeeding events. All the events collectively referred to as an elementary act require an activation energy. This is seen directly from the strong dependence of b on temperature (page 420). The external force applied to each bridge and equal to $P/n = n_0 f/n = f/w$ hinders the transition $1 \rightarrow 2$ and favours the reverse transition $2 \rightarrow 1$. This is directed along the muscle fibre and, hence, at a certain angle ϑ to the bridge (Fig. 12.17). Thus, the bridge is acted on by the force

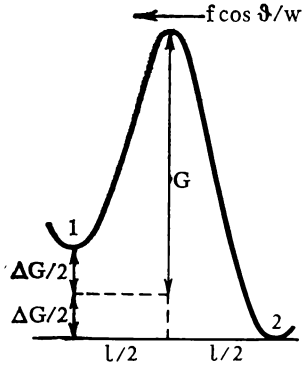


Fig. 12.16. The free-energy curve.

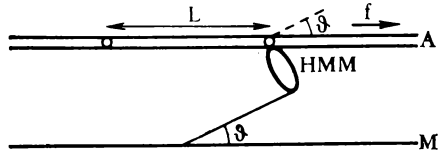


Fig. 12.17. Schematic representation of the operating cross-bridge.

$f \cos \vartheta / w$, to which there corresponds the energy $fl \cos \vartheta / w$, where l is the length of the chemical or chelate bond that links the HMM protruding head to the active centre of actin.

The velocity of shortening will be given by

$$v = (\nu_+ - \nu_-) L \equiv \nu L \tag{12.34}$$

where ν_+ and ν_- are the numbers of elementary acts of transition through the activation barrier per unit time in the directions $1 \rightarrow 2$ and $2 \rightarrow 1$, respectively. According to the Eyring theory we have (see Fig. 12.16):

$$\nu_+ = \frac{kT}{h} \exp\left(-\frac{G_+}{kT}\right), \quad \nu_- = \frac{kT}{h} \exp\left(-\frac{G_-}{kT}\right) \tag{12.35}$$

where h is Planck's constant and k is Boltzmann's constant,

$$\begin{aligned} G_+ &= G + \frac{fl \cos \vartheta}{2w} - \frac{\Delta G}{2} \\ G_- &= G - \frac{fl \cos \vartheta}{2w} + \frac{\Delta G}{2} \end{aligned} \tag{12.36}$$

that is,

$$v = v_+ - v_- = \frac{kT}{h} 2 \sinh \left(\frac{\Delta G}{2kT} - \frac{fl \cos \vartheta}{2wkT} \right) \exp \left(- \frac{G}{kT} \right) \quad (12.37)$$

At $f = f_0$ the velocity of shortening $v = 0$, i.e., $v_+ = v_-$. Here $w = 1$. We obtain $\Delta G = f_0 l \cos \vartheta$ and

$$\vartheta = \frac{kT}{h} 2 \sinh \left[\frac{l \cos \vartheta}{2kT} \left(f_0 - \frac{f}{w} \right) \right] \exp \left(- \frac{G}{kT} \right) \quad (12.38)$$

If

$$\left(f_0 - \frac{f}{w} \right) l \cos \vartheta \ll 2kT \quad (12.39)$$

then it follows that

$$v \approx \frac{l \cos \vartheta}{h} \left(f_0 - \frac{f}{w} \right) \exp \left(- \frac{G}{kT} \right) \quad (12.40)$$

and

$$v \approx \frac{lL \cos \vartheta}{h} \left(f_0 - \frac{f}{w} \right) \exp \left(- \frac{G}{kT} \right) \quad (12.41)$$

We have just derived an equation analogous to (12.27). Comparing Eqs. (12.27) and (12.41), we find that

$$\beta = \frac{h}{Ll \cos \vartheta} \exp \left(\frac{G}{kT} \right) \quad (12.42)$$

and

$$b = \frac{r}{1-r} \frac{f_0 Ll \cos \vartheta}{h} \exp \left(- \frac{G}{kT} \right) \quad (12.43)$$

$$v_{max} = \frac{f_0 Ll \cos \vartheta}{h} \exp \left(- \frac{G}{kT} \right) \quad (12.44)$$

Thus, β, b, and v_{max} are expressed in terms of the molecular parameters f_0, L, l, $\cos \vartheta$, and G.

Let us make numerical estimates. First of all, it is necessary to verify the validity of the condition specified by Eq. (12.39). The tension is maintained in each half sarcomere. For the frog sartorius muscle the number of cross-bridges N in a volume equal to 1 cm^2 multiplied by the length of half sarcomere (1.1 μm) is 6.5×10^{12}. It has been found experimentally that P_0 is of the order of 300 kPa. One bridge produces a force $f_0 = P_0/N = 4.6 \times 10^{-12}$ N. We assume that the length of the linkage is $l = 0.2$ nm and we put $\cos \vartheta < 1$ equal to 0.5. We obtain $f_0 l \cos \vartheta = 4.6 \times 10^{-22}$ J $= 2\ kT \times 0.055$. The condition is fulfilled.

The maximum velocity of shortening of the frog sartorius muscle at 0°C, V_{max}, is equal to 1.33 Λ_0/s, where Λ_0 the initial resting length of the muscle. In our calculations we should take $\Lambda_0 = 1.1$ μm—the half length of the sarcomere. Hence, $v_{max} =$

$= 1.5$ μm/s and $b = 0.4 \times 10^{-5}$ cm/s. The elementary step L has a value of the order of 10 nm. Substituting into Eq. (12.44) the values of f_0, L, $\cos\vartheta$, and l from the experimental value of v_{max} (or b), we find that $G \approx 59$ kJ/mole. When the temperature rises by 10 K (from 300 to 310 K), the value of b must increase by 2.14 times. Experiment gives a 2.05-fold increase. The theory under discussion explains well the temperature dependence of the constant b.

From the equations written above it follows that

$$w = b/(b + v) \tag{12.45}$$

If $v = 0$, then $w = 1$; if $v = v_{max} = b\,(1 - r)/r$, then $w = w_{min} = r$:

$$w_{min} = r = b/(b + v_{max}) \tag{12.46}$$

Naturally, this theory gives the results which are equivalent to those obtained by the Deshcherevsky theory. From Eqs. (12.23) and (12.31) it follows that $k_2/k_1 = (1 - 2r)/r$. Substitution of the steady-state value of $m = nv/Lk_2$ from Eq. (12.20) into Eq. (12.19) yields

$$\dot{n} = k_1 n_0 - \left(k_1 + \frac{k_1 + k_2}{k_2 L}\, v\right) n \tag{12.47}$$

and at $\dot{n} = 0$ we obtain the steady-state ratio $n/n_0 = w$. Equating this expression to relation (12.45), we get

$$k_1 \left(k_1 + \frac{k_1 + k_2}{k_2 L}\, v\right)^{-1} = \frac{b}{b + v} \tag{12.48}$$

From these relations we find

$$\begin{aligned} k_1 &= \frac{l \cos\vartheta}{h} \frac{f_0 r}{1 - 2r} \exp\left(-\frac{G}{kT}\right) \\ k_2 &= \frac{l \cos\vartheta}{h} f_0 \exp\left(-\frac{G}{kT}\right) \end{aligned} \tag{12.49}$$

In the theory under discussion, friction arises as a result of the closing and opening of bridges, since these processes require activation energy. The theory is consistent with the principle of microscopic reversibility and expresses the kinetic constants in terms of the molecular parameters. Nonetheless, this is not a true molecular theory—the detailed molecular mechanism of the contractile process is still unknown.

The theory of heat production by the muscle has been advanced by Tawada, Kounosu, and Oosawa (1974). It is assumed that the energy released during the muscle contraction is made up of the stored energy dissipated at centres of the same type and the energy liberated through the breakdown of ATP at centres of another type. The heat of shortening is produced from the dissipated energy. It is thus presumed that in the two types of centre there act two

different actomyosin ATPases. These suppositions are supported by a number of facts, and theory provides reasonable agreement with experiment. The problem cannot, however, be considered to have been solved, since the theory contains the yet unproved postulates.

12.5. The Kinetic Properties of the Muscle

The quantitative theory expounded above refers to the steady muscle contraction. Of no less interest are the dynamical properties of the muscle under nonsteady-state conditions. The steady contraction has the character of plastic flow. In nonstationary conditions

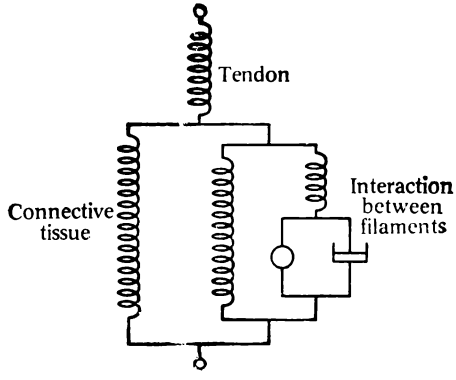


Fig. 12.18. The formal model of the muscle.

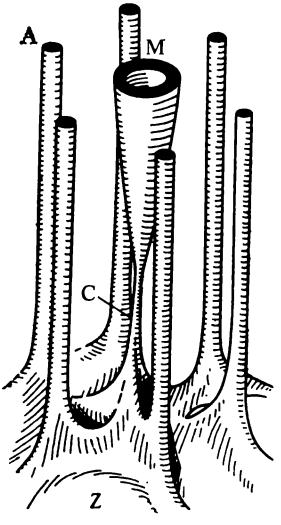


Fig. 12.19. Schematic representation of the microscopic structure of the flight muscle of an insect: *A*—actin; M—myosin.

the muscle exhibits elastic properties. For instance, in quick-release experiments, in which the isometrically contracted muscle is released and experiences rapid isotonic contraction, there are observed slowly damped oscillations. The frequency of these oscillations is of the order of kilohertz for muscles several centimetres long.

If we regard the muscle as a viscoelastic body, we can construct a model containing a damped elastic element connected in series with an undamped elastic element and one more elastic element in parallel with the first two elements (Fig. 12.18). Such a formal

model is a combination of Voigt and Maxwell bodies. The *Voigt model* is an elastic element connected in parallel with a damped element, and the *Maxwell model* contains the same elements connected in series.

The generation of oscillations in the muscle can be determined by the nonlinearity of nonsteady-state kinetic equations that do not contain elasticities in an explicit form. The possibility of oscillations is in this case due to the kinetics of the closing and opening of cross-bridges. On the other hand, the bridge itself is a viscoelastic system. The tension generated by a closed cross-bridge may vary in a stepwise manner, depending on the angle at which the head of HMM S-1 arranges itself relative to actin, and also on the extent of stretching of S-2. The transitions between the steps have an effect on the rapid nonsteady response of the muscle. Thus, the oscillations during quick release are caused by the elastic deformation of the bridge. This is the most probable, though not yet proved, supposition.

Of great interest for physics and biology are the *flight muscles of insects* (IFM) and the related tympanal muscles of cicadas. These muscles are capable of rapid periodic contractions with a frequency of the order of 100 Hz. The flight muscles of insects are structurally similar to the cross-striated muscles of vertebrates. It has been established that the sliding filament model with actin-myosin bridges is applicable to IFM.

The rapid oscillations of IFM require the presence of a directly functioning elastic element. Microscopy shows that, in contrast to the muscles of vertebrates, in IFM there is a direct viscoelastic link between myosin filaments and Z-membranes accomplished by a special element C (Fig. 12.19). Probably, this element is capable of undergoing elastic deformations.

There has been established a large difference between the oscillations of IFM and those of the potential transmitted to these muscles. For example, with flies the frequency of the potential transmitted to IFM is 3 Hz, and the frequency of wing oscillations reaches 120 Hz. Hence, the muscle is not activated with each oscillation. The oscillations of IFM are of the nature of *auto-oscillations*. These auto-oscillations arise in nonlinear systems due to the forces dependent on the state of movement of the system itself; the amplitude of auto-oscillations is independent of the initial conditions (see Chapters 15 and 16). The auto-oscillations of IFM are excited in the presence of a feedback between deformation and stress. The relationship between these varies, depending on the active state of the system. Presumably, IFM contain a "transformer-element" which responds to mechanical events and controls the state of the contractile system. This element is localized in myofibrils, which is proved by the presence of auto-oscillations in IFM preparations washed by glycerol.

During a twitch of the muscle of a vertebrate there is observed a characteristic periodicity of the process, which occurs in some sort of ratchet fashion in finite steps. This was detected by an optical method (Frank, 1964). A very important method of studying the molecular dynamics of the muscle is the rapid X-ray structural analysis based on the application of *synchrotron radiation* (page 150). Vazina and coworkers have succeeded in carrying out an X-ray filming of the muscle with a resolution of up to 0.003 s.

The kinetic properties of the muscle have been studied little. The general approach to the understanding of these properties must be based on the theory of nonlinear dynamical systems (see Chapters 15 and 16).

12.6. Mechanochemical Systems

Muscular contraction is the most thoroughly studied mechanochemical phenomenon. To this category belong numerous biological processes: the movements of plants; the movements of cells with *flagella* and *cilia*; the entire set of motions in *mitosis* and *meiosis*; the *movement of the protoplasm* inside an undividing cell; contractile processes in the tails of phage particles; mechanochemical processes in membranes; the movement of ribosomes relative to mRNA in polysomes; *mechanoreception.*

This incomplete list shows that mechanochemistry is of great importance to biology.

Flagella and cilia are the special contractile systems of bacteria and protozoa that provide their motion in a liquid medium. The cilia function in a number of organs of multicellular organisms. For example, comb-bearers (sea gooseberries) regulate the orientation of their bodies by the coordinated motions of cilia; the cilia create a liquid current in the gills of bivalves, in the human trachea, etc.

Flagella and cilia have a similar structure. They are extended slender filamentous structures with a length ranging from several micrometres to several millimetres and a diameter varying from 0.1 to 0.5 μm. Electron microscopy shows that the flagellum is a bundle of 11 thin fibres, consisting of a circle of nine double fibres around two central single microtubules (Fig. 12.20). The 9 + 2 structural pattern is characteristic of many biological systems that have a fibrillar structure.

The diameter of the central fibres of the flagellum is about 24 nm and the distance between their centres is 30 nm. The lateral size of each of the nine doublets—peripheral fibrils—is 37×25 nm^2. A doublet consists of two subfibrils A and B; projecting from subfibril A are two "arms" 15 nm long and 5 nm thick. The fibrils are hollow rods, whose walls are composed of 10 protofibrils with a diameter of 3.5 nm.

The proteins of flagella and cilia have been studied little. Their ATPase activity has been established. The forms and frequencies of the sine-wave beat propagating along the flagellum from the base to the ending depend on the ATP concentration. The proteins are contractile but they differ from myosin and actin.

The model theories of the beating of flagella and cilia are concerned with changes in the conformational state of protein contractile units when a contractile wave reaches them. In each unit there occurs the splitting of ATP. In flagella and cilia there is sliding between subfibrils and the arms function like the bridges in the actomyosin of the muscle. No detailed molecular theory accounting for the wavelike beating of flagella and cilia has yet been worked out.

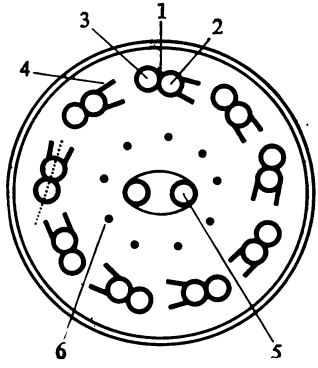


Fig. 12.20. Schematic of the structure of the cross section of the flagellum:

1— peripheral fibril; *2*—subfibril A; *3*—subfibril B; *4*—"arms"; *5*—central fibril; *6*—secondary fibril.

The contractile proteins have also been identified in other systems. From spermatozoa there has been isolated the contractile ATPase protein *spermasin*. The tail structures of bacteriophages, just as motile leaves of plants (say, mimosa) also contain ATPase contractile proteins.

During the motion of the protoplasm in the cell of the seaweed *Nitella* the driving force is generated at the gel-sol interface. There is observed sliding of fibrous structures relative to the surrounding sol. In amoebas there have been detected bundles of thin filaments, and actin-like and myosin-like proteins have been isolated. We may suppose that the actomyosin-ATP system is responsible for the motility of the protoplasm.

During the amoeboid motion in pseudopodia the microtubules are assembled and disassembled.

It has seemed until recently that the wheel was devised by the human mind alone—during the natural evolution there could not have arisen a macroscopic device for rotation about the axis. It has been found, however, that there is something like the wheel even in the bacterium *Escherichia coli*. Each cell of *E. coli* has four long flagella. Their rotational motions account for the motility of the cells. In the basal body of flagella, which resides on the cell wall and membrane, there is a "wheel"—a ring composed of 16 protein molecules in the membrane opposing a similar ring in the cell wall

(this is shown schematically in Fig. 12.21). The rotation of the flagellum caused by the rotation of the ring, which is similar to a ball-bearing, occurs due to the energy released during the transfer of protons into the cell. If one proton is to pass through each protein for the flagellum to be rotated by 360°/16, there are required 256 protons for a complete revolution.

The *plasmodium of myxomycetes* performs active oscillatory motions of the protoplasm. It contains actin and myosin, which are similar

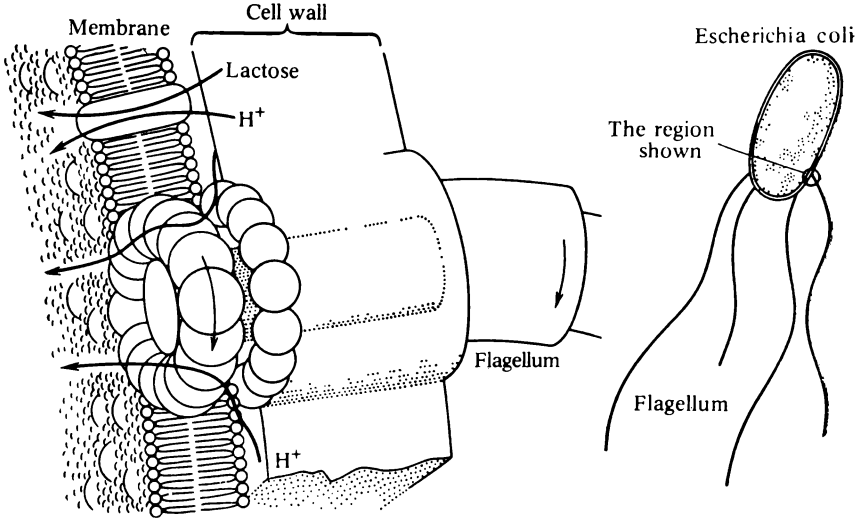


Fig. 12.21. Schematic of the locomotive device of *E. coli*.

to those of the muscle. The myosin of the plasmodium, however, does not form thick filaments; instead it forms only small oligomers. Nevertheless, this myosin interacts with the actin of the plasmodium. In the plasmodium there have been detected thin bundles of filaments composed of actin and myosin. These bundles appear and disappear, following the oscillatory phases. The protoplasm current is regulated by Ca^{2+} ions, probably with the participation of regulatory proteins that have not yet been identified. The motility of the protoplasm might be expected to be similar to the movement of the muscle, but the absence of thick filaments defies this similarity.

In mitosis and meiosis there occurs the movement of the protoplasm and chromosomes. The mitotic apparatus of the cell consists of strands visible under the microscope which connect centrioles with

one another and chromosomes with centrioles. Centrioles have a 9 + 2 structure similar to the structure of flagella and cilia. It has been shown that the flagella of spermatozoa grow out of the centrioles and kinetochore (centromeres) of chromosomes.

The protein of the mitotic apparatus is similar to actin. Injection of ATP causes a slow elongation of the mitotic apparatus. The direct relationship between mitotic motions and ATPase activity has not yet been elucidated, and the mechanism of the mitotic apparatus is still unknown.

There is every ground to believe that *mechanoreception* and also sound reception are associated with mechanochemical processes.

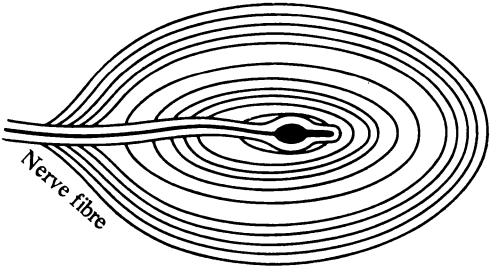


Fig. 12.22. Schematic of the Pacinian body.

It is possible that mechanical reception is tied up with mechanical effects on ionic channels in the membranes of receptor cells, with the deformation of these channels. The ionic channels may be considered to be "vectorial enzymes" that transform the incoming ion (the substrate) to the outcoming ion (the product).

A single mechanoreceptor—the *Pacinian body*—is shown schematically in Fig. 12.22. The receptor consists of a nerve ending bounded by a capsule made up of a peripheral and a central zone. The peripheral zone consists of about 30 closed elastic sheaths. The central zone (the inner bulb) contains 60 tightly packed cytoplasmic layers separated by a cleft oriented along the long axis of the elliptic cross-section of the nerve ending. The modulus of elasticity of the sheaths of the peripheral zone is about 10^6 g/cm^2 (108 Pa); the sheaths are capable of discerning a mechanical force within the range of 150-200 dynes [(1.5-2) 10^{-3} N].

The Pacinian bodies transform a mechanical effect into nerve impulses. In this sense, their function is the reverse of the function of the muscle. Chernigovsky and coworkers (1970) studied the characteristics of these transformers. The maximum sensitivity is

recorded at a frequency of 100-220 Hz; the optimal frequency of the flow of the action potential coincides with this value. A kind of "biomechanical resonance" occurs here. The intrinsic excitation period of the receptor is 9.6 ms, which corresponds to the same periodicity of excitation of the order of 100 Hz. It has been suggested that the regulation of the permeability of the plasma membrane of a nerve ending is associated with the participation of contractile proteins. It has been established that the active site of the nerve ending in the cleft region plays a dominant role in the generation of excitation. It has been demonstrated that ATP is split in this region. One might think that the membrane contains oriented protein molecules exhibiting ATPase activity. A mechanical stimulus causes conformational changes in these molecules and, hence, changes in their ATPase activity.

The most general proposition of biological mechanochemistry is its obligatory relationship with the enzymic activity of working substances—contractile and regulatory proteins. As we have seen (Chapter 6), enzymic activity is determined by the conformational properties of the protein, by electronic-conformational interactions. It follows that the forced conformational change brought about by a mechanical effect on the protein must alter its enzymic activity. This has been proved by experiments. When myosin undergoes deformation in a hydrodynamic field of the dynamooptimeter (page 90), its ATPase activity is changed. The ultrasound has a strong effect on the activity of enzymes.

In principle, similar processes probably occur during sound reception.

Contractile proteins, notably actin, are contained in a number of other cells and tissues. Actin constitutes about 20 percent of the entire protein in the neurons of a chicken. Actin or an actin-like protein is present in the embryonic lens, lungs, skin, heart, the pancreas, kidneys, and in the brain tissue of a chicken. Actin- or myosin-like proteins have been isolated from thrombostenin and from a complex of contractile proteins in the human blood platelets (thrombocytes).

These facts confirm the general proposition concerning the necessity of mechanical motion in life processes and the similarity between the mechanisms of this motion in the large variety of biological systems.

12.7. Biomechanics

Associated directly with mechanochemical processes is the branch of biology and physics known as biomechanics. Here we shall confine ourselves to a brief outline of the problems dealt with in biomechanics.

The movements of vertebrates are controlled by a complex system of muscles, tendons, and bones. The study of these movements requires the solution of a number of problems of the mechanics of a multiply jointed system, hydrodynamics, aerodynamics, and automatic control. The locomotive system of an animal consists of both rigid (the endoskeleton of vertebrates, the chitin exoskeleton of arthropods) and flexible parts. The latter function primarily in the joints. The corresponding systems form *kinematic chains*, i.e., the totality of links so connected that if one of them

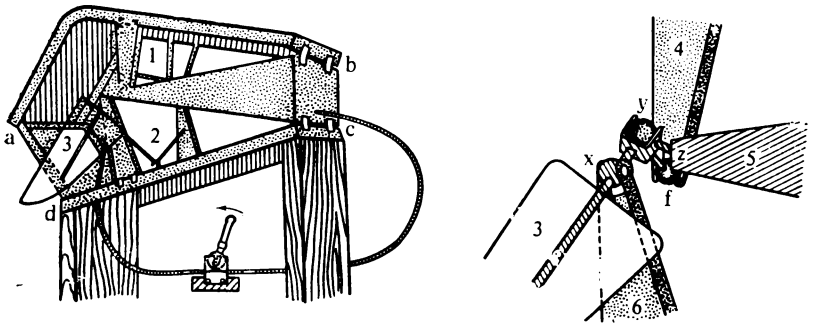


Fig. 12.23. The acting model which explains the mechanism of motion of the wings of a fly:
1—wire depicting the longitudinal spine muscle; *2*—rubber representing the pleirosterial muscle; *3*—wing; *4*—tergite; *5*—shoulder of the scutellum; *6*—pleuritis; *x*, *y*, *z*, *f*—hinges.

is secured and another is set into motion, the remaining ones will of necessity move in a preset manner. The kinematic chains also form limbs and the skulls of the kinetic type in many lizards, snakes, and birds. The kinetic skulls are made of mobile joints.

The task of engineering consists in constructing mechanisms capable of executing definite movements. Biologists are often faced with the inverse problem—to understand the mechanism underlying the motion observed. These mechanisms originated by nature are very complicated, which is evidenced, for example, by the operating model of the wings of a fly (Fig. 12.23).

The mechanical properties of materials and devices employed by organisms for performing motions are varied. The endoskeleton and exoskeleton must be strong enough. For instance, the skull of a woodpecker is built in a special way, allowing strong shocks being absorbed by the bird's braincase. The bones of vertebrates are made of microscopic crystals of *hydroxyapatite*, $3Ca_3(PO_4)_2 \cdot Ca(OH)_2$,

embedded into the fibrillar collagen structure. Collagen provides high bending strength.

Collagen and contractile and other proteins that participate in the movements of animals have specific viscoelastic properties. The arterial walls and other organs of vertebrates contain *elastin*, whose properties are similar to those of rubber. Like rubber, elastin has a high elasticity of the entropy nature (see Sec. 3.1). The breast region of insects contains *resilin*, which also has rubber-like properties—its elastic modulus is close to that of rubber and its elasticity is almost entirely of the entropy nature. During the flight of a locust the greater part of the kinetic energy lost by the wings at the end of their rise is transformed into the potential energy of elastic deformation of resilin; this energy is then converted again into the kinetic energy of the dropping wings. When a flea jumps, its posterior leg, while rising, squeezes a lump of resilin. It ascends relatively slowly as a result of the muscle being contracted and is then abruptly released by means of a special triggering device. Resilin works just as the rubber in a catapult.

The viscoelastic properties of protein materials are very important for movements of both vertebrates and invertebrates.

The movements of animals in air and in water are determined by the evolutionary development of biological systems fitted to perform aerodynamic and hydrodynamic functions that are not, in a number of cases, feasible in artificial devices. The skin of a dolphin has special properties that enable the head resistance to the movement of the animal to be reduced. The waving and gliding flight of birds served as the basis for constructing flying machines by man. The field of science dealing with the technical applications of mechanisms devised by living nature is known as *bionics*. It was founded by Leonardo da Vinci who designed wings for the man to fly. He was also the founder of biomechanics as a whole—he was the first to investigate man's walking, running, and jumping.

Biomechanics also includes the reception of mechanical vibrations and orientation of the body by the organs of equilibrium (the otoliths in the ear).

Biomechanical investigations are highly essential for engineering and medicine, particularly in connection with the weightlessness in space flights. Mention should, for example, be made of an artificial arm set into motion by biocurrents (Gurfinkel). Biomechanics is one of the important branches of biophysics; a detailed consideration of biomechanics, however, is beyond the scope of this book.

Chapter 13 Bioenergetics of the Respiratory Chain

13.1. Biological Oxidation

Bioenergetic processes leading to the synthesis of ATP, the "charging of biological accumulators," take place in *mitochondrial* membranes. Spatially organized molecular systems responsible for the energetics of living organisms are also localized in mitochondrial membranes. The synthesis of ATP in mitochondria is coupled to electron and ion transport and to mechanochemical phenomena. The functions of mitochondrial membranes are very complicated and varied. Another type of bioenergetic coupling membranes—the membranes of plant chloroplasts responsible for photosynthesis—will be discussed in Chapter 14.

The source of the energy required by a living cell for all its various life functions is respiration, i.e., the oxidation of organic compounds by the air oxygen. In 1780 Lavoisier demonstrated that respiration and combustion are of the same nature. The investigations carried out for the last two centuries by chemists, biologists, and physicists have led to the elucidation of the basic features of biological oxidations.

The "fuel", i.e., the substances being oxidized, enters the animal's organism together with food in the form of fats, carbohydrates, and proteins. These are broken down, i.e., are hydrolyzed in reactions catalyzed by special enzymes—fatty acids are activated by means of enzymes and ATP and converted into the acyl derivatives of coenzyme A (page 44). The oxidation of the acyl derivative of CoA-SH involves a number of steps, at each of which there is formed a residue of a fatty acid containing two carbon atoms less than the previous one. The overall equation of the reaction of oxidation of a fatty acid with an even number of carbon atoms to acetyl-S-CoA is as follows:

$$H_3C(CH_2CH_2)_nCO_2H + ATP + (n+1)CoA\text{-}SH + nNAD^+ + nE\cdot FAD + nH_2O \rightarrow$$
$$\rightarrow (n+1)H_3CCOS\text{-}CoA + \begin{pmatrix} ADP + P_i \\ AMP + PP_i \end{pmatrix} + nNADH + nE\cdot Red.\ FAD + nH^+$$

Here NAD is the coenzyme nicotinamide adenine dinucleotide (page 44); E stands for an enzyme; FAD is the coenzyme flavin adenine dinucleotide. The degradation of fats leads, in the long run, to the formation of acetyl-CoA and also of propionyl-CoA and glycerol.

The breakdown and oxidation of carbohydrates (starch, in particular) lead to the formation of triose phosphates and pyruvic acid (pyruvate). When proteins are broken down, apart from individual amino acids, there are formed acetyl-CoA, oxaloacetate, α-ketoglutarate, fumarate, and succinate (succinic acid). These processes of metabolism have been thoroughly studied in modern biochemistry (Sec. 2.10).

The compounds formed undergo further transformations in a cyclic sequence of reactions known as the *citric acid cycle*, the *tricarboxylic acid cycle*, or the *Krebs cycle*. This sequence of reactions is localized in mitochondria.

The Krebs cycle is shown in Fig. 13.1. One revolution of the cycle consisting of eight reactions (numbered in Fig. 13.1) involves the degradation of one molecule of acetyl-CoA or one molecule of pyruvate with the formation of CO_2 and H_2O, i.e., the "burning" of the molecules.

The Krebs cycle starts the pathways of many biosynthetic reactions—the syntheses of carbohydrates, lipids, purines, pyrimidines, and porphyrins. Protein synthesis is also associated with a cycle that produces the precursors of a number of amino acids. At the same time, a biological oxidation is the source of the energy stored in ATP.

The reactions of the cycle give rise to CO_2 and H^+ ions. This is accompanied by the reduction of the coenzymes NAD and FAD. For a continuous and complete biological oxidation to be effected, these coenzymes must be reoxidized. The oxidation is accomplished by a series of electron carriers that form the *respiratory chain* or the *electron-transport chain*. The electron-transport chain ensures the following reactions:

$$3NADH + 1.5O_2 + 3H^+ \rightarrow 3NAD^+ + 3H_2O - 3 \times 220.1 \text{ kJ/mole}$$
$$(\text{Red. FAD}) + 0.5O_2 \rightarrow (\text{FAD}) + H_2O - 152.0 \text{ kJ/mole}$$

The reaction of acetyl-CoA is written as follows:

$$\text{Acetyl-S-CoA} + 3O_2 \rightarrow 3CO_2 + 2H_2O + \text{HS-CoA} - 903.9 \text{ kJ/mole}$$

The electron-transport chain is a polyenzymic system which accepts electrons from the Krebs cycle and the fatty-acid oxidation cycle. The electron is delivered along the chain

$$\text{Substrate} \rightarrow \text{NAD} \rightarrow \text{Flavoprotein}_D \rightarrow \text{Cytochrome } b \rightarrow$$
$$\uparrow$$
$$\text{Succinate} \rightarrow \text{Flavoprotein}_S$$
$$\rightarrow \text{Cytochrome } c_1 \rightarrow \text{Cytochrome } c \rightarrow \text{Cytochromes } a + a_3 \rightarrow O_2$$

Flavoproteins are enzymes containing flavin coenzymes—FAD (Fig. 13.2) and flavin mononucleotide (FMN), riboflavin-5′-phosphate. The cytochromes contain a heme group, the iron atom of which

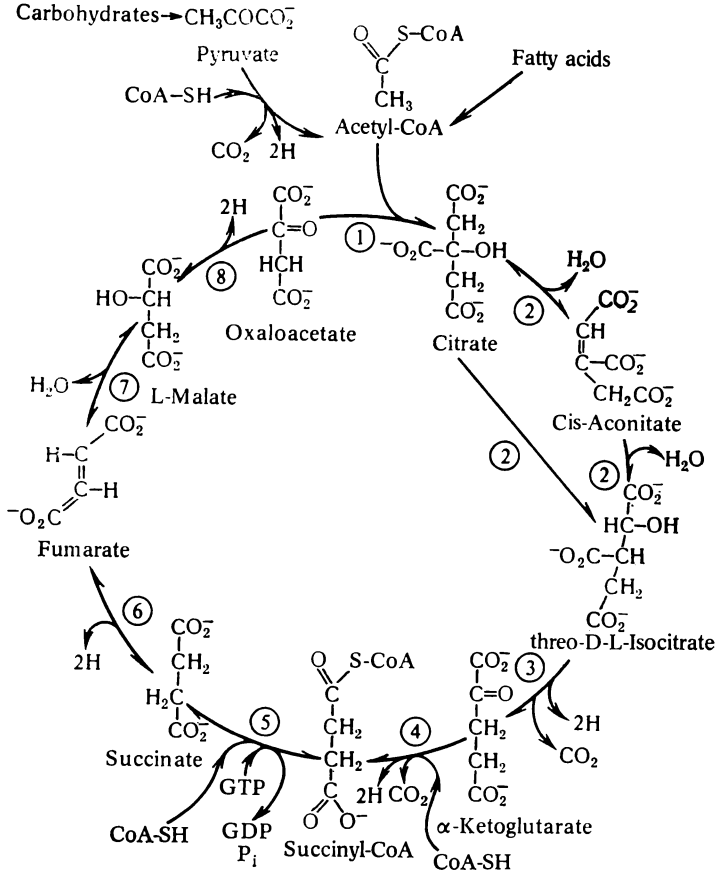


Fig. 13.1. The Krebs cycle (the citric acid cycle).

undergoes oxidation and reduction during the operation of the cycle:

$$Fe^{2+} \rightleftharpoons Fe^{3+} + e^-$$

The electron transfer in the above chain occurs from left to right, terminating in the reduction of oxygen, which unites with H^+, forming water. The electron released on oxidation unites with the

next link of the cycle. The electron transport is accompanied by a free-energy change, since the electrons move through a cascade of increasing oxidation-reduction (redox) potentials. The values of redox potentials are given in Table 13.1.

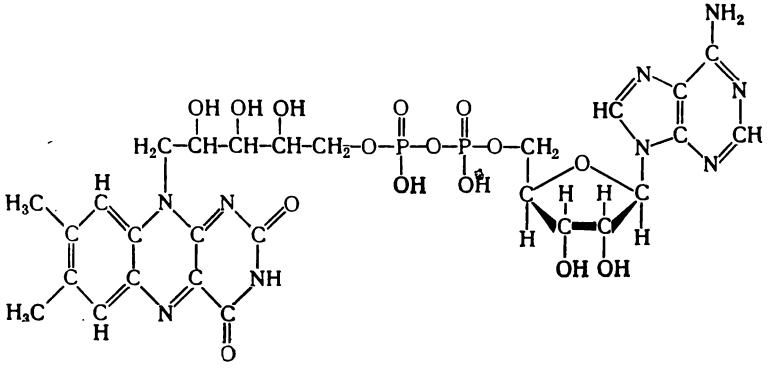


Fig. 13.2. Flavin adenine dinucleotide (FAD).

The electron transport is coupled with the accumulation of energy in the molecules of ATP. In other words, the released free energy is converted to the chemical energy of ATP. Oxidative phosphorylation occurs.

This most important phenomenon was discovered by Engelhardt in 1930. Belitser and Tsybakova (1939) carried out the first determinations of the P/O ratio, i.e., the ratio of the number of molecules of the esterified inorganic phosphate to that of oxygen atoms absorbed, and established that the value of this coefficient is not less than two. It was then found that the P/O ratio for the oxidation reactions of

Table 13.1. Oxidation-Reduction (Redox) Potentials

System	φ, V	System	φ, V
$H_2O/0.5O_2$	0.82	Red. flavin/flavin	—0.185
Cyt. *a* Fe^{2+}/Fe^{3+}	0.29	Lactate/pyruvate	—0.19
Cyt. *c* Fe^{2+}/Fe^{3+}	0.23	Red. NAD/NAD	—0.32
Cyt. c_1 Fe^{2+}/Fe^{3+}	0.21	Malate/pyruvate	—0.33
Cyt. *b* Fe^{2+}/Fe^{3+}	0.05	$0.5H_2/H$	—0.42
Succinate/fumarate	—0.03	Acetaldehyde/acetate	—0.60
Glutamate/α-ketoglutarate	—0.14		

the Krebs cycle and for reactions that involve NAD is equal to 3. Lehninger has found that the processes of oxidative phosphorylation are localized in the mitochondria; he has determined the key steps of the respiratory chain at which phosphorylation takes place. The indicated value of the P/O ratio follows from the equation

$$NADH + H^+ + 3P_i + 3ADP + 0.5O_2 \rightarrow NAD + 4H_2O + 3ATP$$

This equation is the sum total of the exergonic reaction

$$NADH + H^+ + 0.5O_2 \rightarrow NAD^+ + H_2O + 221.3 \text{ kJ/mole}$$

and the endergonic reaction

$$3ADP + 3P_i \rightarrow 3ATP + 3H_2O - 92.0 \text{ kJ/mole}$$

Phosphorylation occurs at three key steps of the cycle: NADH → → flavoprotein, cytochrome b → cytochrome c_1 and cytochrome c → → cytochromes $a + a_3$.

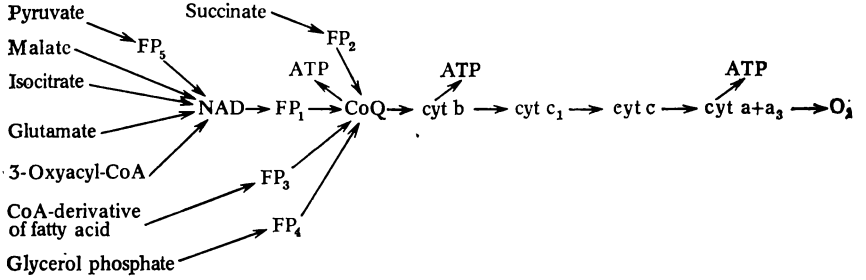


Fig. 13.3. The respiratory chain. The straight arrows indicate the sites of electron entry.
FP—flavoproteins; CoQ—coenzyme Q; cyt—cytochrome.

The general scheme of the coupling of oxidation to phosphorylation is presented in Fig. 13.3.

Let us write again the overall equation:

$$ADP + P_i^- + H^+ \rightleftharpoons ATP + H_2O - \Delta G$$

The free-energy change ΔG is given by

$$\Delta G = \Delta G_0 + RT \ln \frac{[ADP][H_2PO_4^-][H^+]}{[ATP][H_2O]} \qquad (13.1)$$

where

$$\Delta G_0 = -RT \ln K \approx 30.7 \text{ kJ/mole}$$

The quantity K is the equilibrium constant at concentrations of $0.1M$, pH = 7.0, and 25°C. The value of ΔG varies with changes in pH and ionic strength.

The free-energy change during the transport of two electron equivalents along the electron transport chain from NADH to O_2 is determined by the difference of the redox potentials, $0.82 - - (-0.32) = 1.14$ (V), i.e.,

$$\Delta G = z\,\Delta\psi = -2 \times 23.06 \times 4.79 \text{ kJ/mole} = -221.3 \text{ kJ/mole}$$

This provides the synthesis of at least three moles of ATP. The efficiency of the process is equal to 92/211.3 (i.e., about 40 percent). The energy stored in ATP is consumed by the cell for the performance of chemical, electrical, transport, and mechanical work.

The unraveling of the chemistry of a biological oxidation is the most significant achievement of biochemistry.

A specific feature of *oxidative phosphorylation* that distinguishes it from enzymatic reactions in solution is the spatial localization of the stages of the multistep process. Oxidative phosphorylation is localized in mitochondria and is directly associated with the functionality of their membranes.

The oxidation-reduction reactions briefly described above were deciphered as a result of the application of delicate chemical and physical methods. An important role was played, in particular, by the works of Chance which were devoted to the spectroscopy of electron carriers (NAD, FP, cytochromes) in intact mitochondria. These carriers exhibit characteristic absorption bands in the visible and near ultraviolet regions of the spectrum. The difference spectra allow one to study the kinetics of their oxidation and reduction. Various methods were used for the removal of specific enzymes from mitochondria and for retaining only some definite links of the process. The mitochondria were subjected to excision; from the *submitochondrial particles* (SMP) there were isolated complexes of respiratory enzymes free from structural proteins. It was possible to purify such complexes and to study them in detail. Successful experiments were conducted for the restoration of the electron-transport systems from isolated preparations and soluble enzymes. Finally, very valuable information was obtained on the inhibition of individual steps of the process and on the separation of oxidative phosphorylation and electron transport.

Probably, the carriers of electron-transport chains are grouped into four complexes called *Green complexes*. The corresponding scheme is shown in Fig. 13.4. The molecular mass of each of the complexes is about 3×10^5; such a complex is about 64 percent protein and 36 percent lipid. An electron-transport complex is defined as the smallest unit of the electron-transport chain that retains its ability

to transfer an electron at a rate comparable with that in intact mitochondria.

The structure and functional organization of mitochondria were the subject of intensive investigation. Nevertheless, no solution has so far been obtained to many very important questions. The specific membranous structure of mitochondria, the presence of an

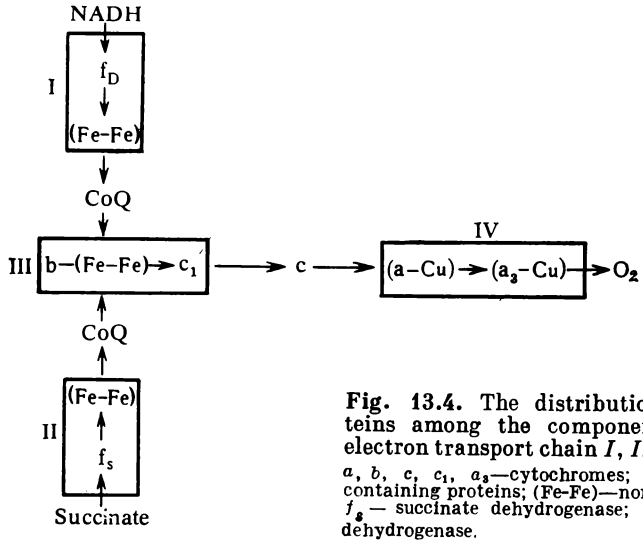


Fig. 13.4. The distribution of proteins among the components of the electron transport chain *I*, *II*, *III*, *IV*: *a*, *b*, *c*, c_1, a_3—cytochromes; Cu—copper-containing proteins; (Fe-Fe)—nonhemin iron; f_S — succinate dehydrogenase; f_D—NADH-dehydrogenase.

autonomic program of protein synthesis (DNA synthesis), and the mechanochemical activity of mitochondria are directly associated with their role of "power plants" of the cell.

13.2. The Structure and Properties of Mitochondria

Mitochondria are present in all aerobic cells of animals and plants, except some primitive bacteria, in which the functions of mitochondria are performed by the plasmatic membrane. The number of such organelles in the cell varies from 20-24 in spermatozoa to 500,000 in the cells of the giant amoeba *Chaos chaos*. The cell of *E. coli* contains one mitochondrion. The number of mitochondria is a characteristic of cells of a given species; presumably, mitosis involves the division of mitochondria and their correct distribution among daughter cells. The shape, structure, and size of mitochondria also vary. They invariably have a system of inner membranes called

cristae. Figure 13.5 is a schematic representation of the structure of the mitochondrion isolated from the rat liver. It is about 3 μm long and 0.5-1 μm wide; it can easily be seen under an optical microscope. The average dry weight is 10^{-13} g. The total surface area of all cristae is 16 μm^2 and that of the outer membrane, 13 μm^2. The cell of the liver contains about 1000 mitochondria and, hence, the total

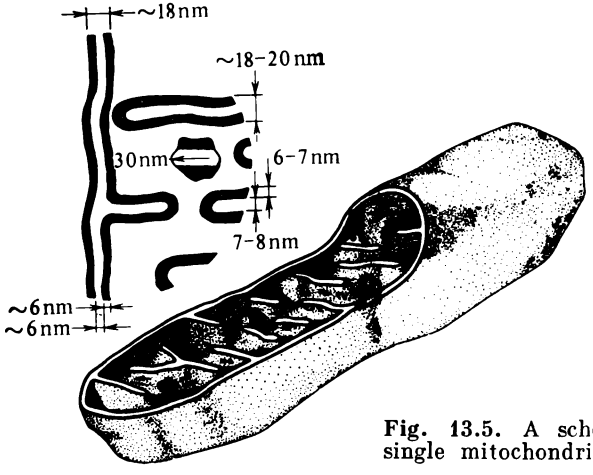


Fig. 13.5. A schematic of a single mitochondrion.

surface area of the mitochondrial membranes is about 29,000 μm^2, which is ten times larger than the membrane surface of the cell itself. The interior of the mitochondrion unoccupied by cristae is filled with a *matrix* material. The mitochondria contain so-called "compact granules" which are apparently formed through the accumulation of various ions, in particular Ca^{2+} ions.

The mitochondria have been found to contain DNA; they are the site of the DNA-dependent synthesis of RNA. The entire system of protein biosynthesis is present in mitochondria—they contain tRNA, aminoacyl-tRNA synthetases, and ribosomes.

Mitochondrial membranes are much thinner than most cell membranes—their thickness is of the order of 5 nm. Electron microscopic examination of mitochondria shows that the inner membrane and cristae are covered by spherical or polyhedral particles 8-10 nm in diameter, which are connected to the membranes through narrow stalk pieces (2-4) by (4-5) nm^2 in size. These structural elements are very numerous, their number reaching 10^4-10^5 in a single mitochondrion; they occupy 10-15 percent of its volume. The headpieces, narrow stalks, and the third unit, a support, constitute what is known

as the elementary particle of the mitochondrion. These elementary particles are seen in Fig. 13.6, which is an electron micrograph of the cristae in the mitochondrion. It is not excluded that these elementary particles are artifacts that result from negative staining.

The mitochondria contain 50-60 enzymes which catalyze the reactions involved in the electron-transport chain, etc. About 40 percent of the entire mitochondrial protein is a structural protein similar to the protein of myelin and bacterial membranes.

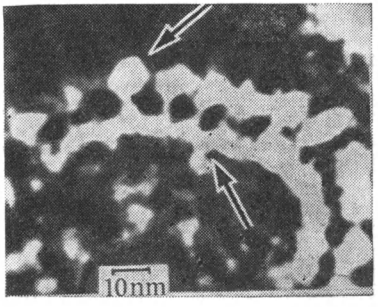


Fig. 13.6. An electron micrograph of the mitochondrial cristae.
The upper arrow indicates the head and the lower the basepiece of the elementary particle.

Lipids constitute from 15 to 50 percent of the dry mass of mitochondria; 90 percent of lipids are phospholipids.

The localization of oxidative enzymes in mitochondria has been studied on mitochondria being destroyed by detergents or ultrasound. The cytochromes and flavoproteins of the respiratory chain are detected in insoluble particles, which apparently belong to membranes. Conversely, the enzymes of the Krebs cycle (dehydrogenases, etc.) are contained in the soluble part. From this we can conclude that they are localized in the vicinity of cristae or directly on them and can easily be separated.

The cristae are formed by the infolding of inner membranes. The number and surface area of cristae have a direct relation to the intensity of cellular respiration. Their shapes vary widely. The mitochondria of the flight muscles of insects have a well-developed surface of tightly packed cristae; the respiration intensity of these cells is 20 times higher than that of the liver cells, whose mitochondria contain a relatively small number of cristae. The mitochondrion of the liver cell is composed of about 17,000 respiratory assemblies, as follows from the number of cytochromes. There are 650 assemblies per every μm^2 of the surface of the inner membrane, each of the assemblies occupying a surface area of 40×40 nm^2.

Phase-contrast microscopy shows that the mitochondria of living cells undergo changes in size and shape associated with respiration. Two types of swelling and contraction cycles occur. A reversible cycle of low amplitude in which the volume is changed by 1-2 percent is observed in all types of mitochondria *in vitro*. Swelling occurs in the absence of ADP in the resting state. When ADP is added, contraction and oxidative phosphorylation of ADP occur. The

cycle is blocked by uncouplers of oxidative phosphorylation.

In the cycle of high amplitude the volume of the mitochondrion can change by several times. This is observed both *in vitro* and *in vivo*. The contraction of mitochondria is brought about by addition of ATP. The process is directly connected with respiration—with oxidative phosphorylation. There is observed a direct parallelism in the inhibition of contraction and oxidative phosphorylation by oligomycin.

These mechanochemical processes are reduced to the conversion of chemical energy to mechanical work. There is a far-reaching similarity between the ATPase activity of mitochondrial membranes and the actomyosin contractile system of skeletal muscles. They have similar mechanochemical properties, i.e., undergo contraction under the action of ATP. One might think that mitochondrial membranes contain contractile proteins that are similar to actomyosin. This hypothesis has been confirmed—a contractile protein has been isolated from mitochondria. It has been shown that contractile proteins participate in mitochondrial mechanochemistry, but it turns out that the membrane lipid phosphatidyl inositol plays a significant role here.

The works of the Lehninger school have demonstrated the direct participation of cation transport in the functioning of mitochondrial membranes. The accumulation of large amounts of Ca^{2+} delivered from the surrounding medium occurs during respiration and is blocked by its inhibitors and by uncouplers of oxidative phosphorylation. The influx of Ca^{2+} is associated with the amount of absorbed oxygen; phosphate is also accumulated simultaneously with the accumulation of Ca^{2+} in the Ca^{2+}/P_i proportion of 1.67 (which corresponds to hydroxyapatite). Electron microscopy shows that granules of calcium phosphate may be accumulated in mitochondria.

There is a direct and precise relationship between the accumulated amounts of P_i and Ca^{2+} and the number of electrons delivered along the electron-transport chain. With transfer of each pair of electrons through each of the three energy accumulation centres there are accumulated 1.67 Ca^{2+} ions and one molecule of P_i. This process is an alternative of oxidative phosphorylation. Thereby, Ca^{2+} acts as its *uncoupler*.

Small amounts of Ca^{2+}, on the contrary, stimulate the respiratory process. The transport of Ca^{2+} is associated with the transport of the cations H^+, K^+, and Na^+.

13.3. Chemiosmotic Coupling

As has been said above, the oxidative phosphorylation in mitochondria is coupled to ion transport. This is a system of coupled processes and not an individual chemical reaction. Slater (1953)

assumed that the oxidation gives rise to *primary macroergs*—intermediate substances having an excess of free energy and taking part in phosphorylation. Such substances have not been detected and the physical mechanism of the process is not clear. The Slater chemical hypothesis has been abandoned.

The *Mitchell chemiosmotic hypothesis* has gained wide recognition (1961). This hypothesis is based on three postulates.

First, phosphorylation occurs only in membranes that contain closed vesicles. The membrane separates the interior of the vesicle from its exterior. It has a low permeability to protons.

Second, it is assumed that electron transfer between the components of the electron-transport chain is accompanied by proton movement across the membrane. The carriers are arranged asymmetrically relative to the two sides of the membrane. The primary macroerg is the *protonmotive force* (pmf):

$$\Delta\tilde{\mu}_H = \tilde{\mu}_H^{entry} - \tilde{\mu}_H^{exit} = RT \ln [H^+]^{entry}/[H^+]^{exit} + \mathcal{F}\Delta\psi = \\ = -2.3RT\Delta pH + \mathcal{F}\Delta\psi \quad (13.2)$$

Third, it is presumed that $\Delta\tilde{\mu}_H$ determines the synthesis of ATP with the aid of a membrane-bound enzyme—*ATP synthetase*. This enzyme acts vectorially, creating a channel for the passage of protons across the membrane. The synthesis of one molecule of ATP requires the passage of many protons.

The enzyme works reversibly. Upon hydrolysis of ATP the protons return back across the membrane, which is energized.

Before considering the Mitchell theory we shall dwell on the general thermodynamic description of chemiosmotic coupling.

For the coupling of two processes—oxidation (o) and phosphorylation (p)—we have (cf. page 322)

$$J_o = L_{11}\frac{\mathcal{A}_o}{T} + L_{12}\frac{\mathcal{A}_p}{T} \\ J_p = L_{12}\frac{\mathcal{A}_o}{T} + L_{22}\frac{\mathcal{A}_p}{T} \quad (13.3)$$

We find

$$\frac{J_p}{J_o} = \frac{q + lx}{l^{-1} + qx} \quad (13.4)$$

where q is the extent of coupling defined as

$$q = \frac{L_{12}}{\sqrt{L_{11}L_{22}}} \quad (13.5)$$

The quantity $l = (L_{22}/L_{11})^{1/2}$ and x is the ratio of the affinity values:

$$x = \mathcal{A}_p/\mathcal{A}_o \quad (13.6)$$

With complete coupling $q = 1$ and $J_p/J_o = l$, independently of x. As we have seen, $P/O = J_p/J_o$, i.e., the number of ATP molecules synthesized through the consumption of one oxygen atom is equal to 3. Hence, $l = 3$. In the absence of coupling $q = 0$ and $J_p/J_o = l^2x = 9\mathcal{A}_p/\mathcal{A}_o$. With incomplete coupling $0 < q < 1$ and the P/O ratio is nonlinearly dependent on x. Experiment shows that the value of q is close to unity.

With proton movement being taken into account, three fluxes must be considered, the third one being the H^+ flow. The dissipation function has the form

$$\sigma = T^{-1}(J_p\mathcal{A}_p + J_o\mathcal{A}_o + J_H\Delta\tilde{\mu}_H) \qquad (13.7)$$

where $\Delta\tilde{\mu}_H$ is given by formula (13.2). The chemical reactions of oxidation and phosphorylation are coupled to the vectorial proton flux—the coefficients L_{OH} and L_{PH} are different from zero.

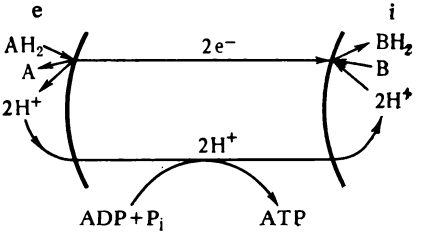


Fig. 13.7. A diagram illustrating the chemiosmotic coupling according to Mitchell:
i—the internal side of the membrane; *e*—the external side of the membrane.

The Mitchell chemiosmotic coupling is shown schematically in Fig. 13.7. The coupling system is the membrane. The hydrogen donor AH_2 (say, ascorbate) is oxidized by an electron carrier (say, cytochrome *c*) on the outer side of the membrane. Two electrons are transferred across the membrane along the respiratory chain and, via cytochromoxidase, are delivered to hydrogen acceptor B, i.e., to oxygen. The acceptor binds two protons from the internal phase of the mitochondrial matrix. This gives rise to a proton concentration gradient—an excess of protons in the external phase and a proton deficiency in the internal liquid phase. As a result of this, protons are translocated across the membrane in the reverse direction, which provides the phosphorylation reaction. The synthesis of one molecule of ATP leads to the absorption of two protons from the external phase and the extrusion of two protons to the matrix. The mitochondrial membrane operates in the same manner as does the *fuel element*, in which an electrochemical potential difference is created as a result of the oxidation-reduction process.

The Mitchell theory relies on the consideration of equilibrium relations. The free-energy difference in the system is made up of chemical and osmotic (transport) works:

$$\Delta G = \Delta G_{chem} + \Delta G_{osm} \qquad (13.8)$$

where

$$\Delta G_{\text{osm}} = \sum_j \nu_j (\tilde{\mu}_j - \mu_j^0) + \sum_i \nu_i (\tilde{\mu}_i - \mu_i^0) \qquad (13.9)$$

The sum over j is the sum over all the products, and that over i is the sum over all the reactants; μ_j^0 and μ_i^0 are the standard chemical potentials: ν_j and ν_i are the stoichiometric coefficients. If

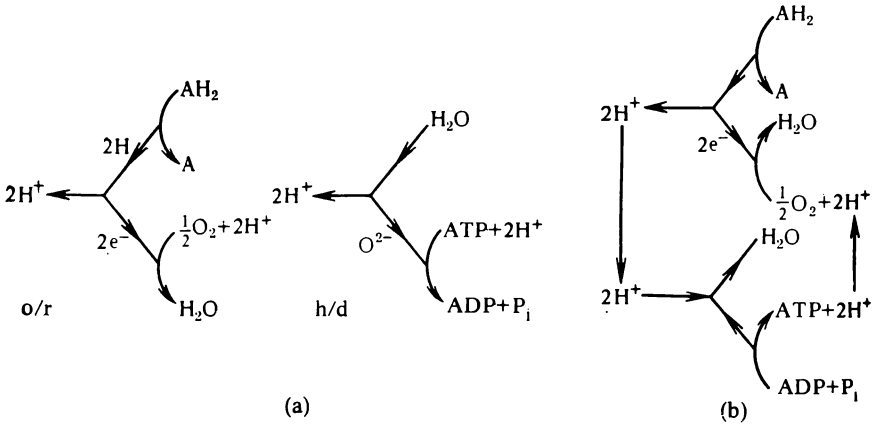


Fig. 13.8. The details of the Mitchell chemiosmotic hypothesis: (*a*) oxidation-reduction (o/r) and hydrogenation-dehydrogenation (h/d) loops; (*b*) coupling between the two processes.

only one component is transferred, the generalized force that acts on it is given by

$$\Delta\tilde{\mu} = -\Delta G_{\text{osm}}/\nu \qquad (13.10)$$

The scheme shown in Fig. 13.3 may be considered to be composed of two loops (Fig. 13.8*a*)—the oxidation-reduction loop (respiration, the *o/r* loop) and the hydrogenation-dehydrogenation loop (dephosphorylation, the *h/d* loop). The coupling of the two processes via proton translocation is shown in Fig. 13.8*b*. According to Eq. (13.10), we have at equilibrium:

$$\Delta\tilde{\mu}_{\text{H}} = -\Delta G_{\text{osm}}/2 = \mathcal{F}\Delta\psi \qquad (13.11)$$

For the *o/r* and *h/d* processes

$$-(\Delta G_{\text{osm}})_{o/r} = -(\Delta G_{\text{osm}})_{h/d} = 2\Delta\tilde{\mu}_{\text{H}} \qquad (13.12)$$

The net electrochemical potential of each component is given by

$$\tilde{\mu}_i = pV_i + z_i\mathcal{F}\psi + \mu_i^0 + RT \ln c_i \qquad (13.13)$$

where p is the pressure; V_i is the molar volume; z_i is the charge of the component; c_i is the concentration (activity). Since the term $V_i \Delta p$ is relatively small, the total generalized force that leads to the ith component in the external phase and in the matrix is

$$\Delta\tilde{\mu}_i \approx z_i \mathcal{F} \Delta\psi + RT\Delta \ln c_i \tag{13.14}$$

For protons, in terms of pH, we obtain formula (13.2):

$$\Delta\tilde{\mu}_H = -2.3RT\Delta\mathrm{pH} + \mathcal{F}\Delta\psi$$

At equilibrium $\Delta\tilde{\mu}_H = 0$, and from Eq. (13.14) it follows that

$$\Delta\psi \equiv \varphi = \frac{RT}{z_i\mathcal{F}} \ln \frac{c_i^{(e)}}{c_i^{(i)}} \tag{13.15}$$

that is, we obtain the familiar Nernst-Donnan formula (page 360).

For the ATP synthesis coupled to the transfer of two electronic charges (this follows from the experimentally found ratio of $P_i/2e^- = 1$ for each centre of phosphorylation) the required value is not less than 200 mV. With a membrane having a thickness of the order of 10 nm the electric field strength in it must not be less than 2×10^5 V/cm. The value of $\varphi = 200$ mV is equivalent to the contribution of $\Delta\mathrm{pH} = 3.5$ to the total protonmotive force.

Mitchell also considered the coupling of fluxes of protons and cations moving in opposite directions (antiport) and the coupling of proton and cation fluxes moving in the same direction (symport).

Thorough determinations of the number of protons, the number of charges, and the number of ATP molecules per each of the three coupling centres in the electron-transport chain have recently been carried out (cf. page 444). With the aid of various electron donors and acceptors there have been obtained the results shown schematically in Fig. 13.9. Each coupling centre has a proton pump which separates 4 charges (4 protons are expelled from the cytoplasm into the matrix)—the H^+/centre ratio is equal to 4 in all the three centres. On passage of two electrons (e^-) through centre I two charges are separated and one ATP molecule is synthesized; in centre II half the ATP molecule is liberated per $2e^-$ and two charges are separated; in centre III there are released 1.5 molecules of ATP per $2e^-$ and 6 charges are separated.

When the ATP synthesis is coupled to the transport of $2e^-$, the required value of φ is not less than 200 mV. If the membrane has a thickness of the order of 10 nm, the electric field strength in it must not be less than 2×10^5 V/cm. The value of φ equal to 200 mV is equivalent to the contribution of $\Delta\mathrm{pH} = 3.5$ to the total proton motive force.

The Mitchell theory has been repeatedly confirmed qualitatively. Liberman and his coworkers have studied ion transport across artificial phospholipid membranes. In the presence of synthetic ions with a charge screened by hydrophobic substituents, say tetrabutylammonium, $N^+[(CH_2)_3\,CH_3]_4$, or tetraphenylborate, $B^-(C_6H_5)_4$, the electric conductivity of the system is substantially increased. These ions diffuse rapidly through membranes. The transport of these ions across mitochondrial membranes (MM) and submitochondrial particles (SMP) produced by treatment of the mitochondria with ultrasound has been studied. The mitochondrial membranes and submitochondrial particles have been found to be oriented in opposite

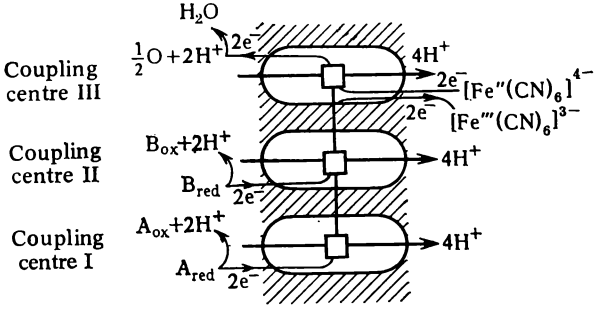


Fig. 13.9. The extrusion of protons and the separation of charges at the three coupling centres.
The squares indicate the unknown mechanisms of coupling of the transfer of e^- to the transfer of H^+.

directions. Cytochrome *c* is localized on the outer side of the MM and on the inner side of the SMP membrane. One might think that the intramitochondrial space is charged negatively and that the inner space of the SMP is charged positively. The energization of submitochondrial particles by addition of ATP causes the influx of synthetic anions, while the de-energization by a respiration inhibitor (actinomycin) or by an uncoupler of oxidative phosphorylation (a derivative of phenylhydrazone) causes the efflux of anions. The electron transport in SMP membranes is accompanied by the influx of synthetic anions. In turn, their transport is upset by electron-transport inhibitors and uncouplers of oxidative phosphorylation.

On the whole experiment shows that an energy-dependent ion transport occurs in coupling membranes. Its mechanism is independent of the structure of ions capable of penetrating phospholipid

membranes. A change in the orientation of membranes relative to the surrounding medium (MM and SMP) leads to a reorientation of ionic fluxes directed against the ion concentration gradients.

According to the Mitchell theory, the transport of protons and electrons across membranes does not lead to considerable changes in pH. Electron transport, i.e., an oxidation process, is stopped under the action of an electric field generated by an excess of negative charges on the other side of the membrane. Electron transport is reactivated with a decrease in the electric field strength, i.e., the membrane potential. This can be achieved through the movement of cations across the membrane. This must give rise to a pH gradient, since the transport of each univalent cation must be accompanied by the extrusion of one proton in the external medium. Addition of permeant synthetic cations to mitochondria causes an increase in the H^+ concentration in the medium. Changes in pH are stopped upon introduction of uncouplers. During the operation of the respiratory chain the medium is acidified. It has thus been established that the energization of a coupling membrane creates a transmembrane potential $\Delta\psi$ and a difference ΔpH.

From the chemiosmotic hypothesis it follows that the respiration energy transformed from the chemical to the electrical and osmotic forms can be reconverted to the chemical form in the ATP synthesis or during the reverse transport of electrons. The bioenergetic process is reversible, which is confirmed by experiment. There is realized an ionic phosphorylation at the expense of the electrical or osmotic energy released upon movement of ions across the membrane down the concentration gradient.

A decrease in $\Delta\psi$ must lead to the uncoupling of oxidation and phosphorylation. This is the result of the action of uncouplers, such as dinitrophenol (DNP) and other substances. Dinitrophenol (2,4-dinitrophenol) presumably acts as a proton carrier. If the Mitchell scheme is correct, then any weak acid or base must exert an uncoupling effect. Uncouplers really increase the proton conductance (permeability) of mitochondrial membranes. There has been revealed a far-reaching similarity between artificial phospholipid membranes and inner mitochondrial membranes.

In the laboratories headed by Liberman and Skulachev, the localization of the respiratory chain was determined from its ability to generate a membrane potential. Various electron donors and acceptors were introduced into the medium which did not pass through the membrane. It was found that these substances interacted only with cytochrome *c* in mitochondria. It was established that the transport of protons and/or electrons along the respiratory chain really takes place. In other experiments there was determined the localization of the components in mitochondrial membranes. Figure 13.10 shows the probable localization of the chain.

According to the chemiosmotic hypothesis, any coupling system must generate an electrochemical potential of H^+ ions. Indeed, experiments with permeant synthetic ions have shown the generation of $\Delta\psi$ in mitochondria, submitochondrial particles, chloroplasts (see Chapter 14), and in bacterial membranes.

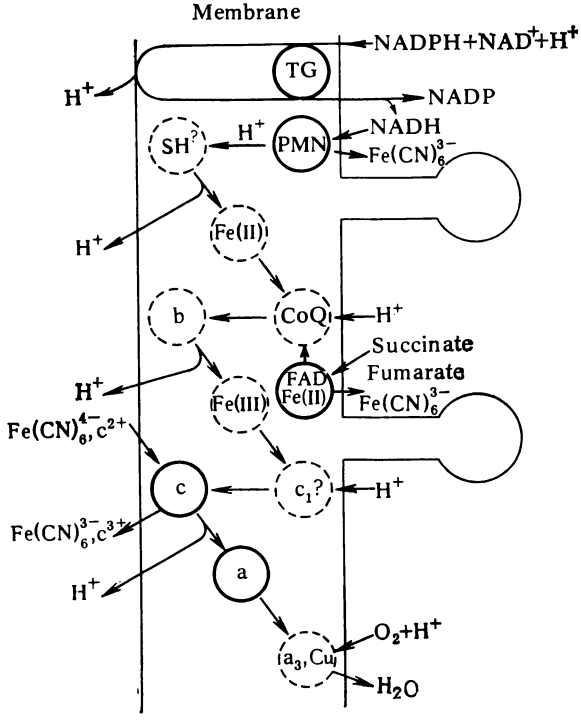


Fig. 13.10. The probable localization of the respiratory chain in the mitochondrial membrane:

TD—transdehydrogenase; a, a_3, c, c_1, b—cytochromes; FMN—flavinmononucleotide; Fe—complexes of nonhemin iron; Cu—copper-containing protein.

At the same time, the Mitchell theory is faced with difficulties and causes objections. Blumenfeld proposes arguments testifying to the impossibility of constructing a "Mitchell machine" in the condensed phase. In such a machine, the ATP synthetase makes use of the difference of proton concentrations in the aqueous phase on the two sides of the membrane for the performance of external work. This is an entropy machine which receives energy from a thermostat in the form of the kinetic energy of protons. Protons move pre-

dominantly down the concentration gradient and transmit their impulses to the mobile parts of the machine; the potential difference $\Delta\psi$ is used up for the generation of "hot" high-energy protons. The corresponding deviation from Boltzmann equilibrium is hardly probable, since the relaxation of excited thermal degrees of freedom occurs rapidly, taking only 10^{-11}-10^{-12} s. Hence, ATP synthetase must be an entropy machine of a different type: the energy received from the thermostat in the interaction of the proton with the synthetase is transformed to a local increase in potential energy. External work can only be done if the elementary act on both sides of the membrane gives rise to a potential difference $\Delta\psi$. ATP synthetase, which interacts with protons, must work as a hydrogen electrode and the entire system of utilization of the proton concentration difference ΔpH must operate as a concentration galvanic cell. The only way of using both components of $\Delta\tilde{\mu}_H$ is to generate a net electric field, and the protons moving inside ATP synthetase must accumulate in the storage battery of unknown nature until an energy sufficient for synthesizing one ATP molecule is accumulated. However, ΔpH cannot be used for the shift of the chemical equilibrium—the entire energy of the nonequilibrium state has been used for the generation of a field.

It also remains unclear how the electron-donor process (the transport of protons down the gradient μ_H) can ensure the accumulation of energy (in ATP) without its dissipation.

Thus, despite the great achievements, the chemiosmotic theory requires further quantitative investigations.

Nevertheless, the importance of the chemiosmotic theory is great. It has served as a powerful impetus for a thorough study of the bioenergetics of coupling membranes and has made it possible to account for a number of facts pertaining to bioenergetics and photobiology (see Sections 14.3, 14.4, and 14.8).

13.4. Electronic-Conformational Interactions

The transfer of electrons in molecular systems is invariably accompanied by the movement of atomic nuclei. As we have seen (Chapter 6), in biopolymeric systems there are realized electronic-conformational interactions (ECI). The movement of electrons and charge densities in a biopolymeric system, in particular in a membrane, coupled with conformational motions may be treated as the movement of a quasi-particle—a *conformon* (page 215). There is every ground to consider oxidative phosphorylation in the electron-transport chain on the basis of these conceptions.

Green and Ji (1972) proposed an electromechanochemical model for the structure and functions of mitochondria. The elementary particles of mitochondria (MEP) are assumed to exist in the ground

unenergized state and in an excited energized state. The free energy of a given state is made up of the chemical, electrical, and mechanical energies. The interconversion of these contributions is determined by electronic-conformational interactions. The concrete Green-Ji model, however, is hypothetical.

Blumenfeld and his coworkers have established the conformational conversions in mitochondrial membranes (with the aid of the EPR spectra of spin-labeled membranes, luminescence spectra, etc.). The high cooperativity of these conversions and the presence of thermal structural transitions in submitochondrial particles have been proved.

Blumenfeld (1972) advanced the hypothesis of *relaxational conformational transitions* in mitochondrial membranes. Let us assume that the elementary act consists of the binding of an electron to an active site (say, to the heme group of the cytochrome). To the neutral form of the macromolecule there corresponds the equilibrium conformation I, and to the charged (reduced) form another equilibrium conformation II. The electron transfer occurs much more rapidly than the conformational change. Hence, during the elementary act there is formed a nonequilibrium conformational state—the electron has been transferred but the conformation has not changed. This is followed by a slow relaxation to state II. The entire transition may be depicted by the following scheme (the minus sign designates an electron):

$$\mathrm{I} \xrightarrow[\text{rapidly}]{} \mathrm{I}^- \xrightarrow[\text{slowly}]{} \mathrm{II}^- \xrightarrow[\text{rapidly}]{} \mathrm{II} \xrightarrow[\text{slowly}]{} \mathrm{I}$$

The energy released in the elementary act is stored in the form of conformational energy during the relaxation time. One might think that this is what constitutes the essence of the primary macroerg and that the hypothesis based on electronic-conformational interactions runs counter to the chemiosmotic theory. During the relaxation time the conformational energy is transformed to chemical energy. It is presumed that at the coupling point there is formed a labile complex of a carrier with a certain group at the active site of the coupling enzyme, the role of which is presumably played by the adenine of the bound ADP. At some moment of the relaxation $\mathrm{I}^- \rightarrow \mathrm{II}^-$ the energy level at which the electron resides drops to the acceptor level of adenine. These two levels are separated by a barrier, but there is possible an under-barrier *tunnelling transition* of the electron to adenine. An increase in electron density on adenine is accompanied by a sharp increase in basicity of the amino group. If there is an electrophilic group (say, a carboxyl group) at the active site of ATP synthetase, the adenine reacts with it to form an amide linkage. At the next relaxation moment the level of the carrier is lowered beneath the level of adenine and the electron density is transferred from adenine back to the redox group of the same or the next electron carrier in the electron-transport chain.

Blumenfeld and Chernavsky (1973) extended this model to any enzymatic reaction. A postulate is formulated, according to which the conformational change of the substrate-enzyme complex that follows the binding of the substrate to the active site of the enzyme involves chemical changes of the substrate in addition to the breaking of the old and re-formation of new secondary bonds in the protein macromolecule. The elementary act of the enzymatic reaction involves the conformational change of the macromolecule (the enzyme-substrate complex, ESC), and the rate of the substrate-product conversion is determined by the rate of this conformational change. The catalytic disruption of the A—B bond of the substrate may be represented by a sequence of four steps:

$$A—B+E \xrightarrow{1} (A—B)E \xrightarrow{2} A\widetilde{E}B \xrightarrow{3} \widetilde{E}+A+B \xrightarrow{4} E+A+B$$

Here step 1 is the rapid step of formation of ESC; step 2 is the slow relaxation of the enzyme E to a new conformation $\widetilde{E}$; step 3 involves the rapid detachment of the product; step 4 is the slow transition $\widetilde{E} \rightarrow E$. All the steps involve a change in free energy.

In order to construct a kinetic theory, one has to determine the relative role of two mechanisms—the over-barrier transition according to Eyring (see Sec. 6.1) and the *tunnel effect*. The latter effect was first considered by Chance in application to electron transfer in a biological system (1968).

The rate constant of the activation transition is given by (see page 191)

$$k_1 = \frac{kT}{h} \exp\left(-\frac{G^+}{kT}\right) \tag{13.16}$$

The number of tunnelling transition acts per unit time is

$$k_2 = \nu_0 \exp\left[-\frac{4\pi L}{h}\sqrt{2m_e(U-E)}\right] \tag{13.17}$$

where ν_0 is the frequency of "vibrations" of the electron, which is of the order of 10^{17} s^{-1}; L is the barrier width; U is the barrier height; E is the energy level of the electron; m_e is the mass of the electron. Figure 13.11 shows the potential-well diagram for two neighbouring links in the electron-transport chain.

The total rate of the transport is determined by the sum of k_1 and k_2. The first term predominates at high temperatures and the second at low temperatures. According to the available estimates, $L \approx$ 1-2 nm, $U \approx G^+ \approx$ 1-2 eV. At these values $k_2 \gg k_1$. From this it follows that the transitions must practically be independent of temperature, which contradicts experiment. This contradiction can be explained.

The tunnelling of an electron depends on the ratio of the electronic level difference ΔE (see Fig. 13.11) and the magnitude of resonance splitting:

$$\Delta E_r = E \exp\left[-\frac{2\pi L}{h}\sqrt{2m_e(U-E)}\right] \tag{13.18}$$

If $\Delta E \gg \Delta E_r$, no tunnelling transition will occur. At $\Delta E \leqslant \Delta E_r$ the tunnelling transition occurs during the time $\tau_r = \hbar\,(\Delta E_r)^{-1}$. At $L = 2$ nm we have $U - E = 2$ eV and $\Delta E_r \approx 10^{-6}$ eV. The values of ΔE in the electron-transport chain must be much larger

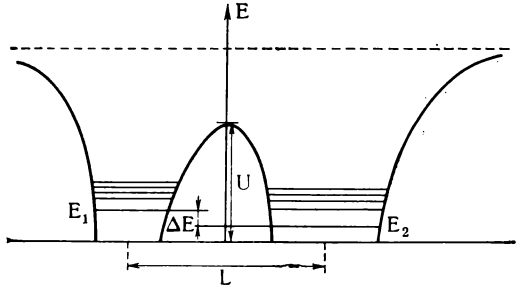


Fig. 13.11. Two asymmetric potential wells separated by an energy barrier.

and such an ideal transition is impossible. Tunnelling occurs, however, because of the dissipation of the energy that provides the approach of the electron levels and which is determined by electronic-conformational interactions. Chernavskaya and Chernavsky consider the potential wells protected from the ionic medium by a nonpolar sheath. Let us assume that only one degree of freedom is essential. There is possible the excitation of a vibrational quantum E_q at which

$$\Delta E = \hbar\omega_q \equiv E_q \tag{13.19}$$

or an increase in the energy of the existing quantum:

$$\Delta E = \Delta E_q \equiv \hbar\Delta\omega_q \tag{13.20}$$

The values of E_q and ΔE_q are of the order of 0.1 eV. The probability of a tunnelling transition on condition (13.19) is expressed by the formula

$$w = \frac{2E^2}{h}\,\rho\,(\Delta E)\,g^2 \exp\left[-\frac{4\pi L}{h}\sqrt{2m_e(U-E)}\right] \tag{13.21}$$

Here $\rho\,(\Delta E)$ is the level density of the final state; $g^2 = 10^{-2}$ -10^{-4} is a constant which characterizes the "bond strength". In the case

of Eq. (13.20)

$$w = \frac{E^2}{\pi\hbar}\rho(\Delta E - \Delta E_q)\exp\left(-\frac{\hbar\omega_q}{kT}\right)\times$$
$$\times\exp\left[-\frac{2L}{\hbar}\sqrt{2m_e(U-E)}\right] \qquad (13.22)$$

The quantity w depends on temperature.

The kinetics of the oxidoreduction of cytochrome c will be discussed in the next section.

In biopolymeric systems a conformon is formed (see page 215). Below we give a crude estimate of its energy and size.

In a medium containing uniformly distributed ions of mass M, which interact in accordance with Coulomb's law, we have (according to the virial theorem):

$$\frac{1}{2}Mv^2 \approx e^2/r \qquad (13.23)$$

To the velocity v and the distance r there corresponds the circular vibrational frequency:

$$\omega_v = \frac{v}{r} = \sqrt{\frac{8\pi}{3}\frac{e^2}{M}n} \qquad (13.24)$$

where

$$n = \left(\frac{4\pi}{3}r^3\right)^{-1}$$

is the number of ions in unit volume. While interacting with the vibrations (conformational transitions) of ions, the electron acquires the energy $\hbar\omega_v$ and the momentum given by

$$p = \sqrt{2m_e\hbar\omega_v} \equiv \sqrt[4]{4m_e^2\hbar^2\omega_v^2}$$

where m_e is the electron mass. According to the uncertainty relation, the corresponding size of the electron cloud is equal to

$$q = \frac{\hbar}{p} = \sqrt{\frac{\hbar}{2m_e\omega_v}} = \sqrt[4]{\frac{3}{32\pi}}\sqrt[4]{\frac{\hbar^2 M}{m_e^2 e^2 n}}$$

Introducing the Bohr radius $a_0 = \hbar/m_e e^2 \approx 0.05$ nm, we obtain

$$q = \sqrt[4]{\frac{3}{32\pi}}\sqrt[4]{\frac{Ma_0}{m_e n}} = a_0\sqrt[4]{\frac{3}{32\pi}}\sqrt[4]{\frac{M}{m_e n a_0^3}} \qquad (13.25)$$

The size of the conformon must be of the same order of magnitude. The polarization energy is equal, in order of magnitude, to

$$E_p = \frac{e^2}{\varepsilon q} \qquad (13.26)$$

where ε is the microscopic dielectric constant. Finally, the ratio of E_p to the vibrational energy gives the number of quanta that

participate in the formation of a conformon:

$$N = \frac{E_p}{\hbar\omega_v} \tag{13.27}$$

For proteins the mean value of charge density is $n \approx 10^{21}$ cm^{-3}, i.e., the separation of the charges is of the order of 1 nm. If the charges reside on protons, then $M = (6.2 \times 10^{23})^{-1}$ g. The value of ε may be considered to be equal to 3. We obtain $\omega_v \approx 2\pi \times 0.6 \times \times 10^{13}$ s^{-1} and the size of the conformon $q \approx 1.5$ nm, i.e., practically the entire molecule is polarized. The polarization energy $E_p \approx \approx 0.35$ eV, the electron level shift being of the same order of magnitude. Finally, $N \approx 15$.

Four processes can be distinguished in the kinetics of electronic-conformational transitions: (1) vibrations of electrons inside the well at frequencies of the order of 10^{17} s^{-1}; (2) vibrations of atoms at frequencies of the order of $kT/h \sim 10^{13}$ s^{-1} and amplitudes of 10^{-2}-10^{-3} nm; the frequencies of elastic vibrations of the protein globule are of the order of 10^{-12}-10^{-11} s^{-1}; (3) electron tunnelling, $\tau \sim 10^{-7}$-10^{-6} s and distances range from 1 to 2 nm; (4) slow conformational transitions in the macromolecule, $\tau > 10^{-3}$ s.

Thus, the rapid processes (1), (2), and (3) may be omitted from consideration; we can average the positions of electrons and atoms with respect to time. It is this procedure that has been employed in the above-described relaxation model. While acquiring an electron, the molecule is found to be in a nonequilibrium conformation, which slowly relaxes to equilibrium. For the tunnel effect to be realized, it is necessary that a suitably arranged virtual electronic level of the acceptor be present in the oxidative conformation and no coincidence of the electron levels of the reduced donor and acceptor is required. The energy liberated on tunnelling is dissipated, but the energy released slowly on conformational relaxation can be converted to the energy of the macroerg. Being associated with the resonance conditions of electron energy levels, the tunnel effect is subject to the influence of the membrane potential. Hence, a regulatory process is possible—the membrane potential generated by the active transport of ions depends on the velocity of electron transport, which is in turn regulated by the membrane potential. A feedback is realized, which makes it possible for the system to be automatically maintained under conditions close to resonance conditions.

The Blumenfeld and Chernavsky hypotheses based on the consideration of electronic-conformational interactions differ substantially from the conceptual models of the Mitchell theory. The concept of electronic-conformational interactions has real and deep-seated physical backing. The nature of oxidative phosphorylation, however, remains unclear. It is obvious that further experimental and theoretical investigations are required.

13.5. Cytochrome *c*

Among electron carriers, the most thoroughly studied one is cytochrome *c*, a universal heme-containing protein and an obligatory participant in the respiration processes of all living organisms, from yeast cells to human beings.

The primary structures of cytochrome *c* of many tens of organisms have been revealed. The cytochromes *c* of vertebrates, insects, lower plants, and higher plants contain, respectively, 104, 107, 107-109, and 111 amino-acid residues. The amino-acid sequence reflects well biological evolution (see also Sec. 17.9). The evolutionary changes of cytochrome *c* in all living things are reduced, as a rule, to replacements of amino-acid residues within a single class (hydrophobic aromatic; hydrophobic aliphatic; hydrophilic basic; hydrophilic acid; ambivalent). This points to a far-reaching similarity between the three-dimensional structures of all cytochromes *c*.

Cytochromes *c* are primeval proteins. During the evolution the cells changed over from anaerobic processes to respiration. It is exactly at this stage, about 1.2×10^9 years ago, that cytochromes *c* originated.

The stability of the three-dimensional structure of cytochrome *c* developed in the course of evolution is proved, in particular, by the fact that the cytochromes *c* isolated from any eukaryote react *in vitro* with the cytochromoxidase of any other species. The structure of cytochrome *c* is shown in Fig. 13.12. The heme group is linked through the fifth and sixth valencies of the iron atom to the imidazole ring of His 18 and to the sulphur atom of Met 80. The heme is also covalently linked to Cys 14 and Cys 17. The immediate environment of the heme—the sequence of amino-acid residues 70-80—is invariant. These residues are the same in nearly all species: -Asn-Pro-Lys-Lys-Tyr-Ile-Pro-Gly-Thr-Lys-Met-.

Naturally, the conformation of cytochrome *c* is believed to be changed upon transitions $Fe^{2+} \rightleftarrows Fe^{3+}$. The first results of X-ray diffraction analysis (Dickerson, Winfield) have shown that when cytochrome *c* is reduced, the positions of several aromatic residues are altered. This direct evidence of electronic-conformational interactions looked very attractive. Unfortunately, these results have not been confirmed. Only recently have there been detected small differences in the three-dimensional structures of the oxidized and reduced forms of cytochrome *c* which boil down to the change in the position of the water molecule that adjoins the heme.

The quantum-mechanical theory of oxidation-reduction conversions of cytochrome *c* has been developed in a number of works, notably in the works of Jortner (1976) and Dogonadze and his coworkers (1977 and earlier works). Jortner considered the temperature dependence of the activation energy for the transport of electrons

between biological molecules. The theory takes direct account of electronic-conformational interactions—many-phonon transitions corresponding to conformational motions are studied. In accordance

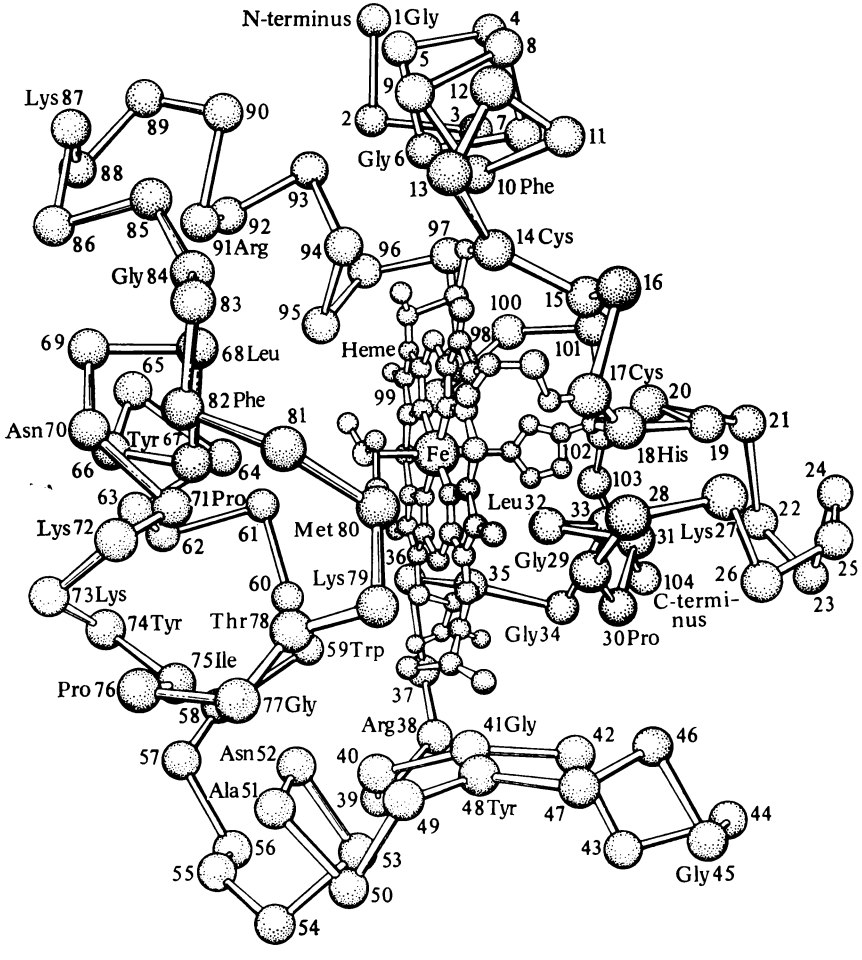


Fig. 13.12. The structure of cytochrome *c*.

with experiment, Jortner obtains, at low temperatures, an activationless transition of electrons and a strong temperature dependence for a transition in a narrow region near $kT/\hbar\langle\omega\rangle \approx 0.1\text{-}0.2$, where $\langle\omega\rangle$ is the mean frequency of vibrations of the nuclei. The temper-

ature dependence of the velocity of electron transfer originates at a temperature much lower than the characteristic vibrational temperature.

Dogonadze, Kuznetsov, and Ulstrup (1977) have developed a rigorous theory of the conformational dynamics of the processes of transport of electrons and atoms in biological reactions. Practical calculations based on this theory are, however, difficult to make—the wavefunctions of the systems are unknown. Of great importance for biophysics are crude qualitative models (pages 211 and 213) and qualitative methods of quantum chemistry (page 216).

It is evident that investigations of oxidation-reduction biological processes are highly topical. Many of the methods here are informative. Particularly, magnetic spectropolarimetry (Sec. 5.8) allows one to directly follow the conversions of various types of cytochromes in the electron-transport chain in a suspension of submitochondrial particles.

Chapter 14 Photobiological Processes

14.1. Photosynthesis

Bioenergetic processes with which we have been concerned in the preceding chapters and which are reduced to the accumulation of the chemical energy of ATP and its consumption are dark processes, i.e., they occur in the absence of light. In this chapter we shall deal with the most important photobiological processes.

Photosynthesis in green plants determines the existence of all higher forms of life, since the oxygen in the Earth's atmosphere was formed exactly as a result of photosynthesis (see Sec. 17.1). The general photosynthetic equation can be written as follows:

$$CO_2 + H_2O + \hbar\omega \rightarrow (CH_2O) + O_2 + 470\ \text{kJ/mole}$$

Here $\hbar\omega$ is a light quantum; (CH_2O) is a fragment of the carbohydrate molecule. The photoreaction liberates an energy of 470 kJ/mole. The free-energy change is 504 kJ/mole and, hence, the change in entropy is equal to -113 J/(mole·K) (at 300 K). For one molecule of O_2 to be formed there are consumed 8 light quanta (see page 467) with a total energy of about 1470 kJ/mole. Thus, the efficiency of utilization of the solar energy is $500/1470 = 0.34$. The liberation of about 470 kJ/mole of energy follows from the following balance:

Energy of two C=O bonds in CO_2	$798 \times 2 = 1596$	
Energy of two O—H bonds in H_2O	$462 \times 2 = 924$	
Total	2520	
Energy of the O=O bond in O_2		487
Energy of two C—H bonds in formaldehyde, CH_2O		$386 \times 2 = 772$
Energy of the C=O bond in CH_2O		798
Total		2057

$$2520 - 2057 = 463\ \text{kJ/mole}$$

The same value is obtained from the oxidation-reduction (redox) potential of the $\frac{1}{2}\ O_2/H_2O$ couple which is equal to $+0.81$ eV and from that of the $CO_2/CH_2O + H_2O$ couple equal to -0.40 eV. The sum total is 1.21 eV. In photosynthesis, for CO_2 to be reduced

to carbohydrate it is required that 4 hydrogen atoms be transferred from H_2O to CO_2:

$$CO_2+2H_2O\rightarrow(CH_4O_2)+O_2\rightarrow(CH_2O)+H_2O+O_2$$

The energy balance is $1.21 \times 4 = 4.84$ eV $\equiv 470$ kJ/mole.

Thus, the photosynthesis in green plants is the reaction between the oxidizing agent CO_2 and the reducing agent H_2O in which 4 electrons (or 4 hydrogen atoms) are transferred "upwards", overcoming a potential difference of about 1.2 V, which is shown schematically in Fig. 14.1. In bacterial photosynthesis, H_2S may be involved in the reaction instead of water; here the potential difference of H_2S/S is only 0.2 V. The energy required for the reaction is furnished by light. The first step of the process is the absorption of light by pigments, the most important of which is *chlorophyll*. Timiryazev showed that the spectrum of action of the solar light in photosynthesis corresponds to the absorption spectrum of chlorophyll. Figure 14.2 presents the structural formulas of the most important pigments of plants; the absorption spectra of the pigments are given in Fig. 14.3.

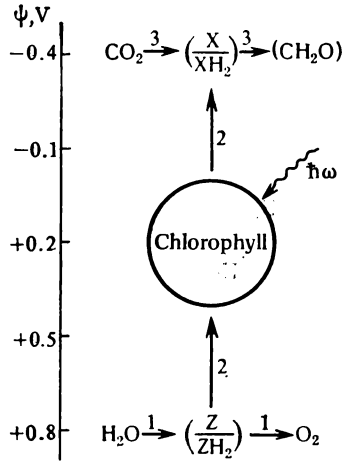


Fig. 14.1. Schematic representation of photosynthesis:
1—enzymatic conversion of H_2O to O_2; 2—transfer of hydrogen from the intermediate substance Z/ZH_2 to X/XH_2 in the sequence of enzymatic steps with the aid of light-activated chlorophyll; *3*—enzymatic conversion of CO_2 to (CH_2O).

The rate of photosynthesis measured by the amount of O_2 liberated per unit time (or of the CO_2 absorbed) depends on the intensity of the incident light, I. Roughly speaking, the rate of formation of a certain substrate is proportional to the number of absorbed quanta. This unstable substrate undergoes further transformations in enzymatic processes. Experiment shows that for one molecule of O_2 to be produced $n \approx 8$ substrate molecules are required. There are about 300 chlorophyll molecules (50 in photosynthesizing bacteria) per one enzymic complex or one molecule of the generalized enzyme (the *photosynthetic unit*).

The rate of photosynthesis may be given by a formula similar to the Michaelis-Menten equation (page 194):

$$v = \frac{(k_E/n)\,EI}{K+I} \qquad (14.1)$$

where k_E is the rate constant of the enzymatic reaction; $n \approx 8$; E is the enzyme concentration. At a high incident light intensity, $I \gg K$, there occurs saturation, $v_{max} = k_E E/n$. The value of k_E is found by measuring the dependence of the O_2 yield on the time interval between flashes, t_d, on pulse illumination. The yield of O_2

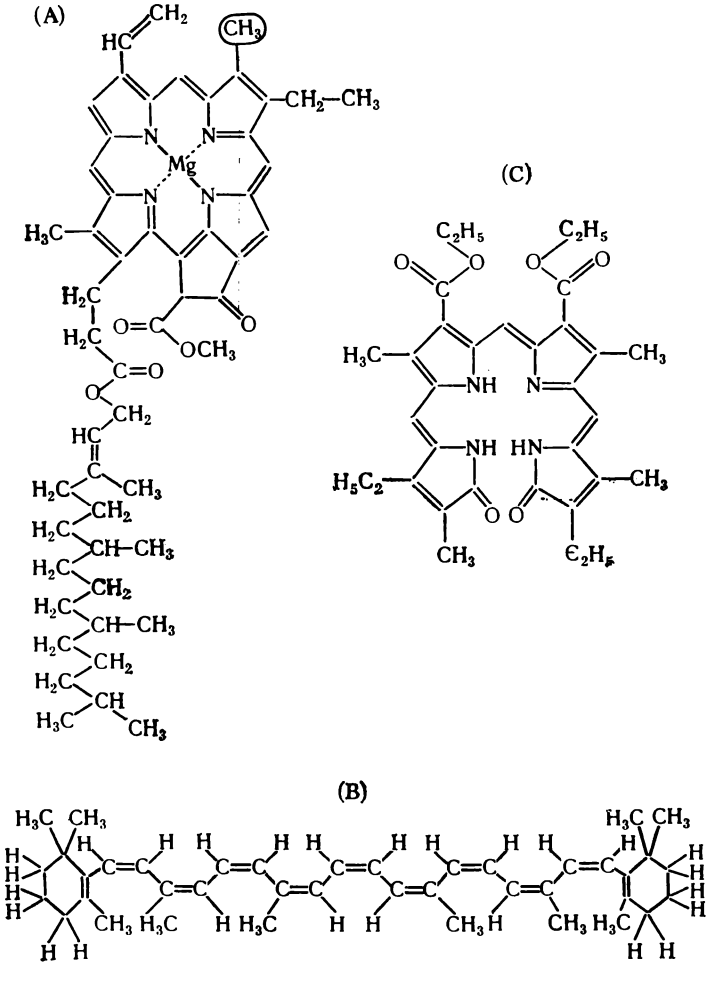


Fig. 14.2. The structural formulas of some pigments:
(*a*) chlorophyll *a*; in chlorophyll *b* the CH_3 group enclosed in a circle is replaced by —HC =O; (*b*) β-carotene; (*c*) phycoerythrobilin.

per flash increases with t_d up to $t_d \approx 0.1$ s. At this value saturation takes place. Hence, the rate-limiting enzyme can practically process the entire amount of substrate in 0.1 s. The average time required for the conversion of one substrate molecule is about 0.02 s. In other words, $k_E = 1/0.02 = 50$ (s^{-1}). The concentration E = [Chl]/300,

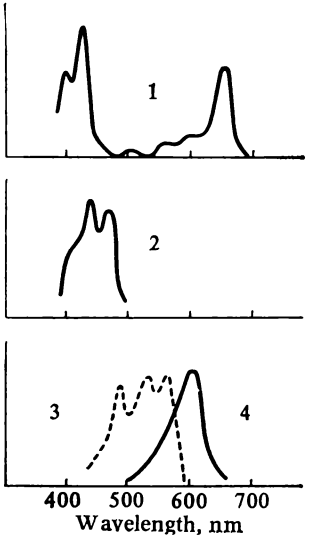


Fig. 14.3. The absorption bands of chlorophyll *a* (*1*), carotenoid (*2*), phycoerythrin (*3*), and phycocyanin (*4*). The ordinate represents optical density.

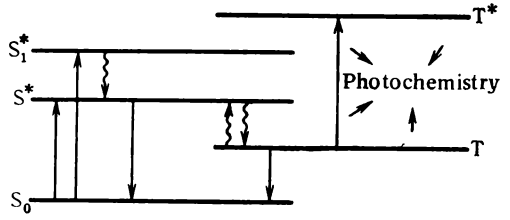


Fig. 14.4. The electron level diagram for chlorophyll according to Krasnovsky.

where [Chl] is the concentration of chlorophyll. We obtain $v_{max} = 50$ [Chl]/$n \cdot 300 \approx 0.02$ [Chl] s^{-1}. The maximum rate of photosynthesis with constant illumination is equal to one O_2 molecule per chlorophyll molecule in 50 s.

In fact, however, the photosynthesis involves a series of enzymatic reactions. With flashes of 10^{-3} s the maximum yield of the process increases and becomes dependent on temperature.

The photosynthetic units are localized in *chloroplasts*, which are special organized cell organelles (see Sec. 14.4). In water suspensions of chloroplasts with ferric salts the *Hill reaction* takes place, which is the photochemical production of O_2 involving the added oxid nt. Apart from Fe^{3+}, quinones and dyes may also serve as oxidants.The oxidant replaces the system $CO_2/(CH_2O)$. The source of O_2 in photosynthesis *in vivo* and *in vitro* is H_2O and not CO_2, which has been proved by means of labeled ^{18}O atoms.

The photosynthetic process is accompanied by respiration, i.e., the absorption of O_2. The difference of these two processes is mea-

sured in the presence of light. The rate of respiration is independent of illumination.

The key problem of the physics and chemistry of photosynthesis is to elucidate the mechanism of conversion of light energy to chemical energy. In 1948 there was discovered the *Krasnovsky reaction*—the reversible photor duction of chlorophyll:

$$\mathrm{Chl} + \mathrm{AH} \overset{\hbar\omega}{\rightleftharpoons} \mathrm{Chl}^- + \mathrm{AH}^+$$

The reducing agent, i.e., the electron donor, may be the ascorbate AH. The reduced chlorophyll can be reoxidized. Probably, the chlorophyll molecules excited by light are reduced *in vivo* at the expense of water. The electronic levels of chlorophyll are shown in Fig. 14.4. In the dark the molecule resides at the lower singlet level S_0. The absorption bands of Chl correspond to the transitions $S_0 \rightarrow S^*$ (in the red region of the spectrum) and $S_0 \rightarrow S_1^*$ (in the blue region). The reverse transition $S_1^* \rightarrow S^*$ is radiationless, and the transition $S^* \rightarrow S_0$ is a fluorescent transition. A radiationless transition from the S^* level to the metastable triplet level T is also possible. Upon light absorption in state T chlorophyll passes over to the excited triplet state T^*. Terenin has carried out a general examination of the role of triplet states in the photochemistry of dyes. It is determined by a long lifetime and chemical unsaturation due to the presence of unpaired electrons. The reversible photoreduction of chlorophyll may be represented by the following scheme:

$$\mathrm{Chl} + \hbar\omega \rightarrow \mathrm{Chl}^* \rightarrow \cdot\mathrm{Chl}\cdot(\mathrm{T})$$
$$\cdot\mathrm{Chl}\cdot + \mathrm{AH} \rightleftharpoons \cdot\mathrm{Chl}^- + \cdot\mathrm{AH}^+$$

The dots signify unpaired electrons. The formation of free radicals during the course of the photoreduction of chlorophyll has been proved by electron paramagnetic resonance.

The absorption spectrum of Chl *a* has a complex band in the red region. This makes us believe that there exist several forms of chlorophyll. The red absorption band is shifted when the plant's leaves grown in the dark become gradually green in the light. The presence of several forms of chlorophyll is proved by a change in the absorption spectrum upon gradual extraction of chlorophyll from the leaves.

The light energy absorbed by chlorophyll and other pigments can accumulate, migrate from the molecule of one pigment to that of another and between identical molecules, be emitted (fluorescence and phosphorescence), and dissipated, being converted into heat. In all cases the primary photoact consists of light absorption. Presumably, one of the functions of accessory pigments is to supply energy to chlorophyll *a*.

The *action spectrum* is defined as the dependence of the rate of production of oxygen, P, on the number of absorbed quanta, I_a,

and the wavelength λ, i.e., the function

$$f(\lambda) = P/I_a \qquad (14.2)$$

If P is the number of molecules of O_2 produced in 1 s and I_a is the number of quanta absorbed in 1 s, then $f(\lambda) \equiv \Phi(\lambda)$ is the quantum yield of photosynthesis. When one molecule of O_2 is produced, 8 quanta are absorbed. The $\Phi(\lambda)$ curve for chlorella is shown in Fig. 14.5. There is observed a characteristic "*red drop*"—a sharp decrease in $\Phi(\lambda)$ in the region of 680 nm. It is also observed in the action spectrum of the Hill reaction.

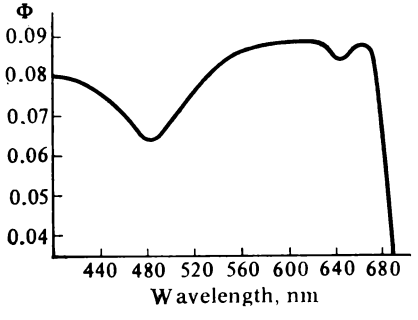


Fig. 14.5. The action spectrum of chlorella.

The red band of Chl *a* is composed of two components, 670 and 680 nm, of approximately equal intensity. There is evidence pointing to a third weak component at 695 nm. Perhaps, these components correspond to different forms of Chl *a*, which differ in photochemical efficiency, this accounting for the red drop.

The actual explanation is more complicated. There exists what is called the *Emerson effect*, which consists in the following: light that is inefficient in the far red region becomes efficient upon simultaneous illumination of the system with shorter-wavelength light, say, upon combined illumination with light at $\lambda = 700$ and 650 nm. The action of double illumination is sharply inadditive. The Emerson effect is expressed by the quantity

$$\varepsilon = (x - y)/z$$

where x is the rate of O_2 production upon double illumination; y is the same upon short-wavelength illumination; and z upon long-wavelength illumination. The value of ε is especially high when the greater part of the additional long-wavelength light is absorbed by the accessory pigment. For maximum efficiency of photosynthesis it is required that one quantum be absorbed by chlorophyll *a* and another by the accessory pigment. A red drop occurs in the region in which light is absorbed by chlorophyll *a* alone. A detailed study has shown, however, that in addition to the $\varepsilon(\lambda)$ maxima corresponding to the accessory pigments, the absorption band of chlorophyll *a* also contains a maximum at 680 nm. This maximum is associated with the resonance transfer of excitation energy from other pigments to the active form of chlorophyll *a*.

14.2. Two Photochemical Systems

The facts described above lead to the hypothesis of two photochemical systems operating in photosynthesis, photosystem I (PSI) and photosystem II (PSII). In green plants photosystem II contains chlorophyll *a* with absorption maxima at 670 and 680 nm. Photosystem I contains all the three forms of chlorophyll *a*. Photosystem II fluoresces strongly, whereas photosystem I fluoresces weakly. The quanta absorbed by 670- and 680-nm chlorophyll are transferred in photosystem II to chlorophyll a_{695} and their energy is dissipated, which is what accounts for the low fluorescence yield. Photosystem II contains more chlorophyll a_{670} than does photosystem I. The light is absorbed in both photosystems simultaneously but to different extents. At $\lambda > 650$ nm in red algae and at $\lambda > 680$ nm in green algae photosystem I receives an excess energy, while photosystem II receives an insufficient energy—the balance required for the joint operation of the two photosystems is upset. This balance is improved upon simultaneous illumination by light of a shorter wavelength. This accounts for the Emerson effect and the red drop.

If the light reactions governed by photosystems I and II occur consecutively, the absorption spectrum of each of the systems can be established by measuring the action spectra of photosynthesis upon strong, though not saturating, action of the light absorbed by the other system. The rate of the overall reaction will then be determined by a limiting process. If the excess light is absorbed by photosystem I, then the reaction rate and the action spectrum will be limited by photosystem II, and vice versa. In this way there have been established the absorption spectra of both photosystems (Fig. 14.6). Photosystem I contains more of 680- and 695-nm chlorophyll *a* and photosystem II contains more of chlorophyll a_{670} and chlorophyll *b*. One cannot say that chlorophyll a_{670} sensitizes one reaction and chlorophyll a_{680} the other. The two reactions are interconnected by a relatively slow dark reaction. The Emerson effect is also observed when two mutually perpendicular beams of light of differing wavelengths enter the reaction vessel. Under such conditions the short- and long-wavelength light is absorbed by the cells of algae at different times. The Emerson effect manifests itself upon interrupted action of light of different wavelength at an interval of several seconds between flashes. In the red algae *Porphyridium* the release of O_2 is enhanced with a flash of green light if it is preceded by a flash of red light absorbed by chlorophyll *a* in photosystem I. It means that a long-lived intermediate compound is formed. Its half-time of decay is estimated at 18 s. Conversely, the production of O_2 with a red light flash is not enhanced if it is preceded by a green light flash; the green light absorbed by photosystem II does not give rise to long-lived intermediate compounds.

Figure 14.7 shows a scheme of photosynthesis which makes clearer the scheme given in Fig. 14.1.

How are the absorbed quanta distributed between the two photosystems, so that they operate at the same rates? Probably, photosystems I and II are spatially separated, which is evidenced by experiments involving the fractionation of the substance of chloroplasts. The balanced excitation is governed by the presence in both systems

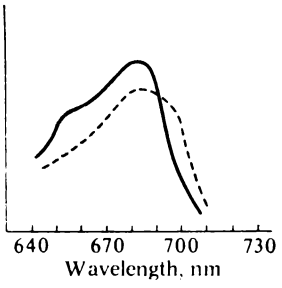


Fig. 14.6. The absorption spectra of PS I (the dotted line) and PS II (the solid line) in chlorella. The ordinate indicates optical density.

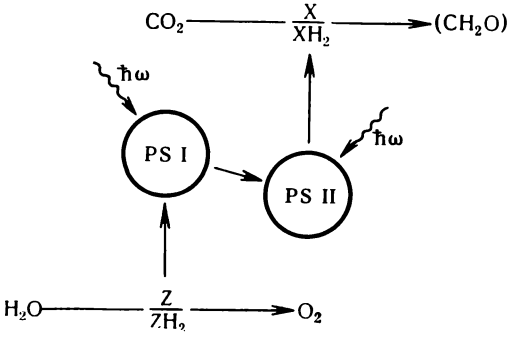


Fig. 14.7. Schematic representation of the photosynthetic process involving two steps sensibilized by PS I and PS II.

of the same pigments, though in different proportions. The quantum yield of photosynthesis must have minima in regions in which one of the systems absorbs light more strongly than the other. There is observed a corresponding fine structure of the action spectrum.

The study of differential absorption spectra has furnished extensive information on the two photosystems. The differential spectra are examined upon simultaneous absorption of light that provides photosynthesis. The exciting light beam is directed perpendicular to the analytical light. The changes in the differential absorption of weak light are studied upon illumination with a strong monochromatic light (in particular, flash light) as compared with the dark. By varying the excitation wavelength one can obtain an action spectrum and determine the pigments that sensitize the changes in absorption. This method was devised by Duysens (1956) and the flash technique was developed in the works of Kock (1963) and Witt (1959). The flash technique allows one to measure the lifetimes of the intermediate compounds and the times of their appearance and disappearance within the interval from 10^{-1} to 10^{-8} s.

The participation in photosynthesis of carotenoid, which is converted upon illumination into a metastable form, has been established by means of these methods. With extra illumination the energy is dissipated through this metastable form. The carotenoids protect chlorophyll against photo-oxidation with an excess of light. The photo-oxidation of chlorophyll a_I (Chl *a* in photosystem I) is governed by the transition of an electron from chlorophyll a_I to $NADP^+$. The oxidation of chlorophyll a_I and the reduction of $NADP^+$ have been found to occur in a 1 : 1 stoichiometry. The final electron donor for chlorophyll a_I is water.

Light with $\lambda < 700$ nm causes reversible changes in the absorption of chlorophyll a_I. The oxidation is followed by reduction during 20 ms. In the region of 700-730 nm chlorophyll a_I is oxidized irreversibly. The pigment in photosystem II is excited only at $\lambda < 700$ nm. Presumably, this is also chlorophyll *a* (chlorophyll a_{II}). The difference between chlorophyll a_I and chlorophyll a_{II} is determined by the difference in their environment. The lifetime of chlorophyll a_{II} is two orders less than that of chlorophyll a_I. The properties of these chlorophylls are given in Table 14.1.

Table 14.1. The Properties of Chlorophylls I and II (after Witt)

Property	Chlorophyll a_I	Cholrophyll a_{II}
Type of reaction	Electron donor	Sensitizer (?)
Absorption bands, nm	438, 660, 682-703	435, 640, 682
Appearance time, ns	$\leqslant 20$	$\leqslant 20$
Lifetime (20°C), ms	20	0.2
Region of excitation	< 730	< 700
Deactivation temperature, °C	65	55

The interplay between photosystem I and photosystem II is realized by means of *plastoquinones* (PQ) that take part in oxidation-reduction (redox) reactions. The excited chlorophyll a_{II}^* reduces plastoquinone to hydroquinone; chlorophyll a_I^* reduces it again. The plastoquinones form a pool containing about 5 molecules between the two photosystems. The intermediate position between the two photosystems is also taken by cytochromes, primarily by cytochrome *f*. Figure 14.8 shows the energy diagram of events that occur in photosystems I and II; and Fig. 14.9 is the general scheme of photosynthesis.

Valuable information about the two photochemical systems is provided by a study of fluorescence. The quantum fluorescence yield of chlorophyll *a* (see page 472) is 30 percent *in vitro* and only 3-6 percent *in vivo*. In the second case several forms of Chl *a*, which differ

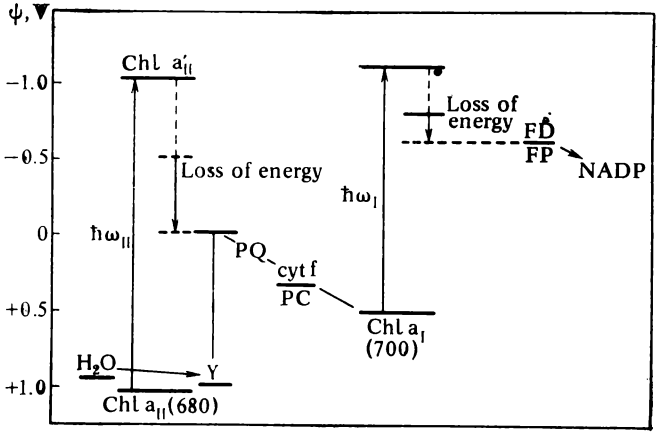


Fig. 14.8. The energy diagram of events that occur in PS I and PS II: PC—plastocyanin; PQ—plastoquinone; FD—ferredoxin; FP—flavoprotein.

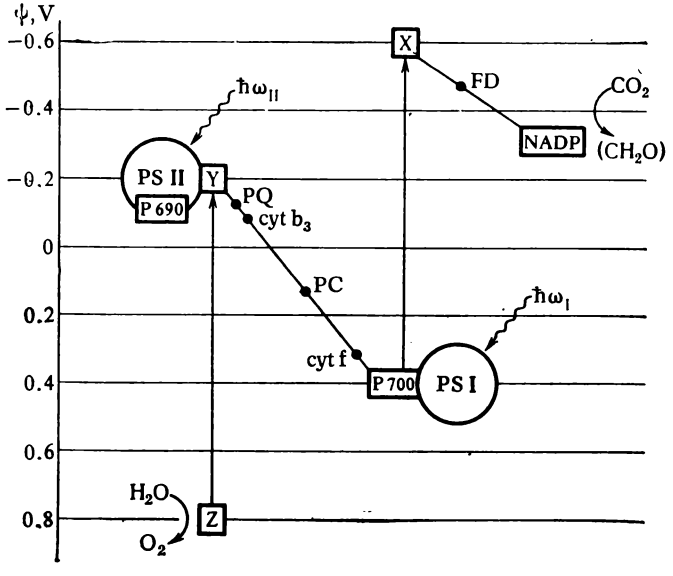


Fig. 14.9. Diagrammatic representation of photosynthesis:
Z and Y are primary donors and acceptors of electrons of light reaction II; P700 and X are primary donors and acceptors of reaction I; P690 = reaction centre of PS II; P700 = reaction centre of PS I; PQ = plastoquinone; PC = plastocyanin; FD = ferredoxin. The ordinate represents the redox potentials in volts.

in emission spectra and quantum yields, fluoresce. The low quantum yield *in vivo* is ascribed to the fact that the energy is consumed for photosynthesis.

The main products of the two primary photochemical reactions are the reduced intermediate compound XH and the oxidized compound Z (see Fig. 14.1). The compound Z oxidizes H_2O to O_2. The compound XH reduces CO_2. These processes are enzymatic. The enzymic system that participates in the evolution of O_2 contains manganese. The Mn^{2+} ion takes part in the reaction photocatalyzed by photosystem II. The following reaction of the photo-oxidized Mn^{2+} is likely to occur:

$$2Mn^{3+} + H_2O \rightarrow \frac{1}{2}O_2 + 2Mn^{2+} + 2H^+$$

for which energy greater than +0.75 eV is needed.

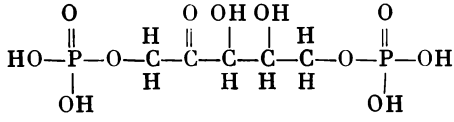
Evidently, Z is an enzyme with the cofactor Mn^{3+}. It was assumed that X is $NADP^+$ (the $NADP^+$/NADPH couple has a potential of −0.35 V). In fact, there are at least two precursors—the iron-containing protein ferredoxin (FD) and the enzyme ferredoxin-NADP-reductase. The illuminated chloroplasts reduce the pigment up to −0.60 V. The potential of X is presumably close to this value. The scheme of possible reactions is as follows:

$$\begin{array}{ccccc} \underset{\text{XH}}{\text{X}} & \rightarrow & \underset{\text{Red.FD}}{\text{FD}} & \xrightarrow{\text{FD-NADP-reductase}} & \underset{\text{NADPH}}{\text{NADP}^+} \\ -0{,}6\text{ V} & & -0{,}4\text{ V} & & -0{,}35\text{ V} \end{array}$$

The pathway leading from CO_2 to carbohydrate is known better owing to the works of Calvin and his coworkers, who used a radioactive ^{14}C label. The equation that describes the fixation of CO_2 has the form

$$6RDP + 6CO_2 + 18ATP + 12NADPH \rightarrow 6RDP + HP + 18ADP + 17P_i + 12NADP$$

Here RDP is ribulose-1,5-diphosphate



HP is hexose phosphate; RDP is the primary acceptor of CO_2. The carbon pathway leading to photosynthesis—a sequence of dark reaction—is described by the *Calvin cycle*, which involves 13 steps. Given above is the overall reaction. In abbreviated notation it looks as follows:

$$6C_5 + 6C_1 \rightarrow 6C_5 + C_6$$

where the numerical subscripts signify the number of carbon atoms in the molecule. The simplified scheme of the Calvin cycle is present-

ed in Fig. 14.10. The hexose phosphate formed is hydrolyzed with the elimination of the phosphate. For one hexose molecule to be formed from 6 molecules of CO_2 the energy of 18 molecules of ATP is consumed. These molecules are formed as a result of the transport

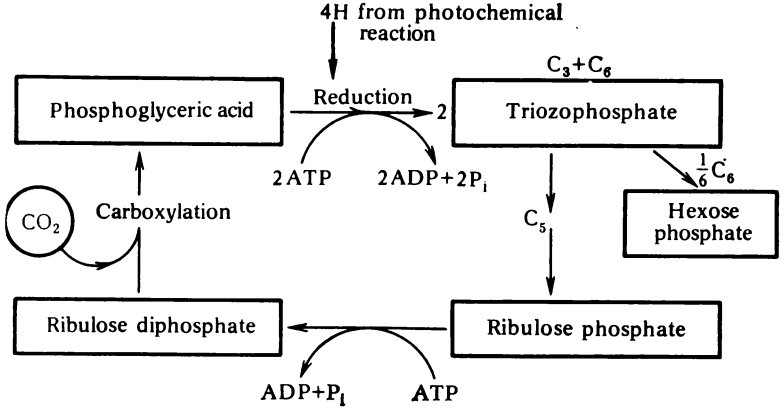


Fig. 14.10. A simplified diagram of the Calvin cycle.

of electrons during photosynthesis—ADP is phosphorylated. The overall reaction is

$$NADP+n'ADP+n'P_i+n\hbar\omega \rightarrow NADPH+n'ATP+O_2+(n'-2)H_2O$$

This reaction is accomplished in the complete system involving pigment systems and the electron-transport chain.

14.3. Chloroplasts

Photosynthesis takes place in the organelles of plant cells called *chloroplasts*. Figure 14.11 shows an electron micrograph of a section of chloroplast isolated from a corn leaf. It is 3-10 μm in diameter and 1.5-3 μm thick. The chloroplast fills almost the entire cell of a green algae. Figure 14.11 shows nearly parallel *lamella* embedded in a lighter stroma. In higher plants the lamella form stacks called *grana*. Lamella are the cross-sections of flattened closed sacs known as *thylakoid disks* having a diameter of about 500 nm. Their number in chloroplast is of the order of 1000. The structural model of chloroplast is shown in Fig. 14.12. Photosynthetic processes are localized in the thylakoid membranes which contain active pigments, primarily chlorophyll. Thylakoid fragments accomplish the Hill reaction and photophosphorylation.

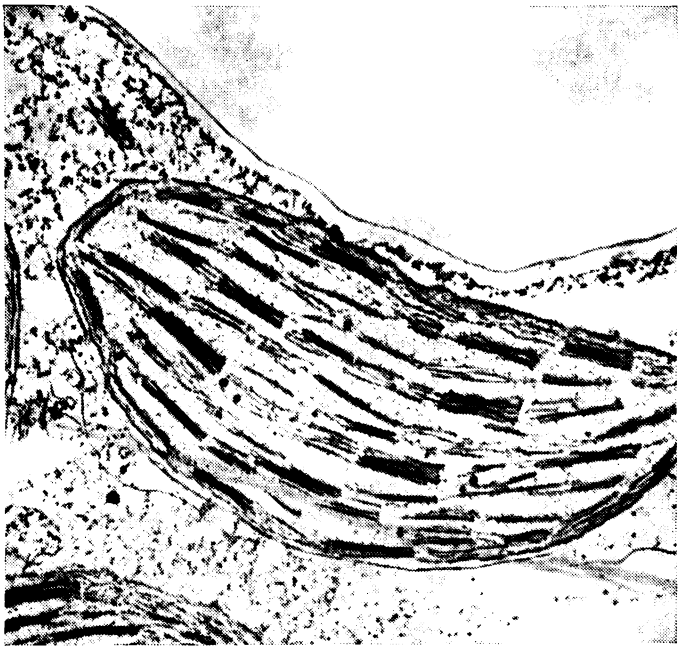


Fig. 14.11. Electron micrograph of a section of chloroplast.

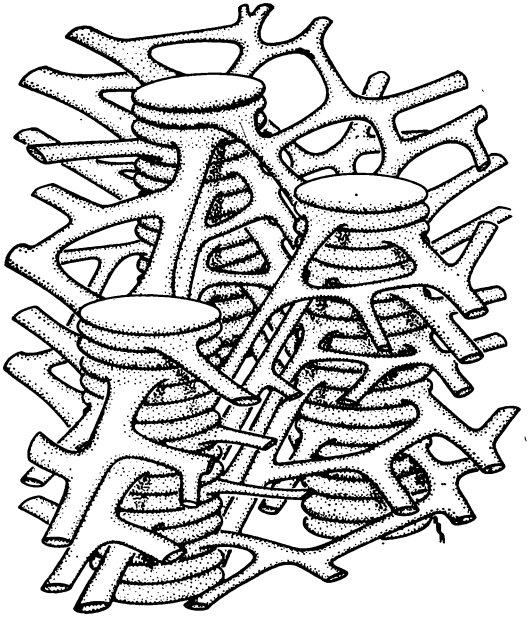


Fig. 14.12. A model of the structure of chloroplast.

The study of thylakoids by means of electron microscopy shows that they contain regular crystal-like systems of tightly packed granules. The granules in spinach thylakoids are 18.5 × 15.5 × × 10 nm^3 in size. The molecular mass of the granules is estimated at 2×10^6. Knowing the composition of the lamella, we can determine the number of molecules in various substances present in the granule. The pertinent data are given in Table 14.2.

Table 14.2. The Composition of Granules from the Spinach Thylakoid

Lipid fraction (number of molecules)	Molecular mass	Protein fraction	Molecular mass
230 chlorophylls	206,400	Proteins	928,000
48 carotenoids	27,400	2 manganese atoms	110
48 quinones	31,800	12 iron atoms	672
116 phospholipids	90,800	6 copper atoms	218
144 digalactosyl diglycerides	134,000		
366 monogalactosyl diglycerides	268,000		
48 sulpholipids	41,000		
Steroids	15,000		
Unidentified lipids	175,000		
Sum total:	989,400	Sum total:	929,000

The number of chlorophyll molecules equal to 230 is consistent with an independent estimate for the photosynthetic unit (see page 467).

The granules are presumably built up of organized subunits. It has been found that chloroplasts contain two species of particles: light particles 7-11 nm in diameter and heavy particles having a diameter of 15-18 nm. The light particles contain more of chlorophyll *a* and are capable of cyclical phosphorylation and of reduction of $NADP^+$ in the presence of ascorbate. The heavy particles contain more of chlorophyll *b* and can reduce ferricyanide in the light (the Hill reaction) and accomplish a noncyclical phosphorylation. The light particles are identified with photosystem I and the heavy ones with photosystem II. There are about 4000 light particles and about 2000 heavy particles per 1 μm^2 of thylakoid surface area. But, as we have seen, the number of photosystems I and II must be approximately the same. It follows that each heavy particle contains two photosystems II.

In operating chloroplasts there are observed gradual changes in light absorption at 475 and 515 nm, which take seconds. In the same spectral regions there occur rapid changes in light absorption,

the bands appearing during the time less than 10^{-5} s and disappearing in 20 ms. The rapid changes are indicators of redox reactions that occur concurrently with phosphorylation. They reflect changes in the electric field across the membrane and their disappearance is associated with the transport of protons and other ions. Indeed, as the ionic permeability of the membrane increases as a result of an osmotic shock these changes disappear more rapidly. This also occurs under the action of uncouplers of phosphorylation, which increase the proton permeability and also under the action of ionophores (gramicidin D).

Experiments have been carried out with layers of chlorophylls *a* and *b* and carotenoids on glass plates. The layers were exposed to the action of electric fields having a strength of up to 10^6 V/cm and the changes in light absorption were measured. They were found to be identical with the light-induced changes in chloroplasts. The change in absorption $\Delta A = \Delta I/I$ is linearly dependent on the change in potential $\Delta\psi$:

$$\Delta\psi = q/C = b\Delta A \tag{14.3}$$

Here q is the charge transferred across the membrane; $C \approx 1$ μF is its capacity; $b \approx 50$ mV/ΔA. The current is given by

$$i = C\frac{d\Delta\psi}{dt} = Cb\frac{d\Delta A}{dt} \tag{14.4}$$

With a flash causing one revolution, two elementary charges are transferred across the membrane per one electron-transport chain. In such a case, $\Delta\psi \approx 50$ mV, with long flashes $\Delta\psi_{max} \approx 200$ mV; with constant illumination $\Delta\psi \approx 100$ mV. An electric field is generated half by the reaction in photosystem I and half by the reaction in photosystem II.

Changes in pH that accompany photosynthesis have also been established. The value of pH increases with decreasing electric field strength. The ratio of the number of absorbed protons to that of electrons transported is equal to 2. These results are in qualitative agreement with the Mitchell chemiosmotic theory (see Sec. 13.3), according to which the transport of electrons provides a free energy for phosphorylation in the form of ψ and pH gradients.

With a short flash $\Delta\psi \approx 50$ mV but the value of pH is less than unity. Under these conditions photosynthesis takes place independently of ΔpH. It has been shown that with a single flash the phosphorylation really occurs due to $\Delta\psi$ with an ATP yield of 0.30 per flash and independently of the time interval between flashes. The phosphorylation is blocked by gramicidin D. At $\Delta\psi > 50$ mV the amount of ATP formed is proportional to $\Delta\psi$. The electrical energy is used for the production of ATP.

14.4. The Mechanism of Photosynthesis

Figure 14.13 shows the general scheme of the coupling of various processes taking place in thylakoid membranes during photosynthesis. This scheme was proposed by Witt (1971).

The events numbered in Fig. 14.13 are as follows:

1. The migration and dissipation of energy. The energy is stored and then dissipated in the excited carotenoid Car* and also as a result of fluorescence.
2. The light reactions in photosystems I and II.
3. The generation of an electric field as a result of a directed electron transport at right angles to the membrane which is governed by the anisotropic orientation of the molecules contained in the membrane.
4. Electron transport and the influx of protons into the membrane in accordance with the Mitchell theory: two electrons on the outer side of the membrane are replaced by two OH^- ions, and two holes (+) on the inner side by two protons. The possible mechanism involves proteolytic reactions coupled to electron transport from H_2O to $NADP^+$.
5. The generation of ΔpH. The bound OH^- and H^+ can be replaced by Cl^- and K^+. The OH^- and H^+ ions diffuse, respectively, into the external and internal phases, thus creating ΔpH.
6. The field-controlled efflux of H^+ and the decrease in ΔpH.
7. The field-controlled efflux of K^+.
8. The diffusion-controlled decrease in ΔpH.
9. Phosphorylation coupled to the discharge of the electrically energized membrane in accordance with the Mitchell theory.
10. The regulation of the inner value of pH^i. At $pH^i \approx 8$ the potassium conductance of the membrane is greater than the H^+ conductance. During the generation of H^+ (process 4) the increase in $[H^+]^i$ slows down the electron flux, which in turn retards the generation of H^+. The increase in $[H^+]^i$ enhances the efflux of H^+, which occurs due to the efflux of K^+. The concentration $[H^+]^i$ is maintained at a certain level. It increases with increasing flash frequency or light intensity. Processes 6 and 9 are also accelerated.

Thus, the energy of light quanta charges the membrane. The right part of Fig. 14.13 shows how this energy is used.

Photosynthesis is a complex series of photo- and electrochemical processes, whose detailed mechanisms are still unknown in many respects. The Witt scheme is so far hypothetical. In quantitative estimations the application of the Mitchell theory to photosynthesis is beset with difficulties. In this theory the free energy of protons must exceed the energy of phosphorylation:

$$n_H G_H \geqslant G_{ph} \tag{14.5}$$

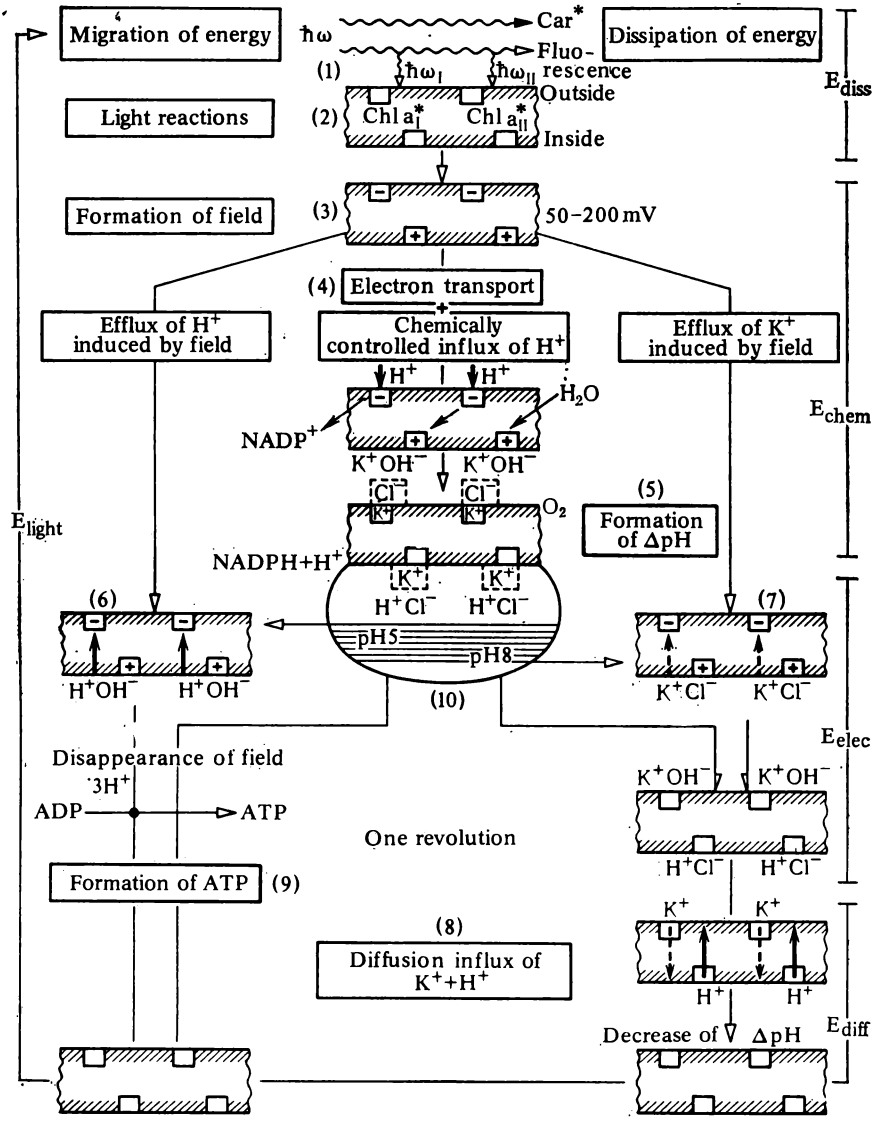


Fig. 14.13. Diagram showing the coupling of molecular processes involved in photosynthesis according to Witt.

Here $G_{ph} \approx 0.5$ eV = 48.3 kJ/mole; n_H is the number of protons transferred. Under natural conditions at $\Delta\psi = 50$ mV and $\Delta pH \approx \approx 2\text{-}2.5$

$$G_H = e\Delta\psi + kT \cdot 2.3\Delta pH \approx 0.18 \text{ eV} \qquad (14.6)$$

Hence, n_H must not be less than 3. Experiment gives a close estimate—for 1 mole of ATP to be synthesized there are required 4 protons. On the other hand, from the ratios $ATP/2e^- = 2$ and $H^+/e^- = 2$ it follows that 2 protons are consumed for the synthesis of one molecule of ATP. Their energy is equal to 0.36 eV, i.e., less than that required for ATP synthesis, and inequality (14.5) is not fulfilled—the energy of one proton is insufficient. This discrepancy reveals itself more distinctly under the conditions of pulse illumination. Further investigations are needed for the Mitchell theory to be applied to phosphorylation in photosynthesis.

Let us consider, in conclusion, the elementary processes that control photosynthesis. Their classification has been proposed by Borisov (1973). The following four classes of phenomena are to be considered:

(a) The absorption of light by chlorophyll and accessory pigments.

(b) The migration of energy from light-absorbing molecules to reaction centres.

(c) The primary separation of charges of different sign.

(d) The stabilization of new energy carriers for coupling with the subsequent chemical steps.

The first group of phenomena have been thoroughly studied and described above. Let us now turn to the second group. It has been established that energy migration really exists and occurs by a resonance mechanism. Other mechanisms of energy migration are also possible, in principle. In the semiconductor mechanism, the excitation of a molecule by light transfers an electron into the conduction zone, after which a pair of opposite charges (an electron and a hole) is displaced in an ordered molecular system, in a molecular crystal. However, in chloroplasts strict order is characteristic of only small assemblies of chlorophyll molecules. It is for this reason that the semiconductor mechanism is hardly probable. At the same time, molecular order in chloroplasts is sufficient for the migration of energy by other mechanisms. There is possible a migration of the energy of triplet excited states or a resonance migration of the exciton or inductive type. In the last two cases, a dipole-dipole interaction leads to the transfer of the energy of a singlet excited state to an unexcited molecule. In exciton migration it occurs during time periods comparable with periods of vibration—it takes 10^{-12} to 10^{-15} s to distances of up to 1.5-1.8 nm. This is the most rapid mechanism. The inductive-type migration is preceded by the attainment of a thermal equilibrium of the excited molecule with the medium,

and the rate of migration strongly depends on temperature. A number of facts are in favour of the rapid exciton mechanism.

The primary physical steps of photosynthesis, beginning with absorption of a quantum and ending with its capture in the reaction centre, occur during the time not longer than about 100 ns. This time is 1.5-2 orders less than the characteristic times of fluorescence of dye molecules; it is the maximum time for diffusion-controlled reactions. Such a rapid energy migration provides a high (more than 90 percent) quantum yield of the primary photosynthetic process.

The problem of the nature of the components of the electron-transport chain has not yet been solved. Various schemes of the electron-transport chain have been proposed. An important specific feature of the electron-transport chain has been elucidated, namely, that the rate of electron transport is independent of temperature at low temperatures. This has led to the natural assumption of the tunnelling mechanism of electron transport between neighbouring carriers. This conception has been developed in the works of Blumenfeld and Chernavsky (see Sec. 13.4).

Electron transport occurs between the centres fixed in the thylakoid membrane. These centres may be treated as sufficiently deep electron traps. One may visualize such a centre as a potential well and consider the levels that the electron being transferred occupies. Here one deals with the problem of electron transport between two neighbouring components of the chain, i.e., between the principal levels of potential wells separated by a barrier (Sec. 13.4). In such a case, in a photosynthetic system there occur electronic-conformational interactions, which can conveniently be treated by resorting to the potential-well model (see Sec. 6.6). After the electron-transport has taken place, the acceptor molecule is found to be in a nonequilibrium conformation, which is slowly relaxing to equilibrium. This makes possible a balanced resonance and a tunnel effect. No strict quantitative theory has, however, been worked so far.

14.5. Vision

Apart from photosynthesis, another most important photobiological process is *photoreception*, which consists in receiving information about the factors of the outside world by means of light radiation. Let us consider the most perfect form of photoreception—the vision of vertebrates.

The structure of the eye is shown schematically in Fig. 14.14. The eye is a self-regulating system. Its optical shortcomings are to a considerable extent compensated by regulatory mechanisms that optimize the operation of the eye. The most important of these are: the focusing of an image on the retina and the regulation of the amount of light that falls on the retina.

The focusing or sharpening-up of the image on the retina is accomplished through automatic change of the radius of the lens curvature, i.e., via accommodation. The shape of the crystalline lens is controlled by a ring of muscles surrounding the lens; these are called the ciliary muscles. The contraction or relaxing of the ciliary muscles

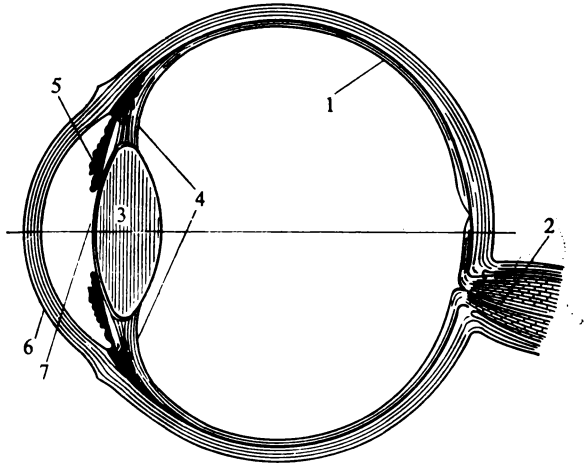


Fig. 14.14. A cross-section through the mammalian eye:
1—retina; *2*—optic nerve; *3*—lens; *4*—ciliary muscle; *5*—iris; *6*—conjunctiva; *7*—pupil.

occurs in response to the defocusing of the image. The accommodation system is a scanning system, since it brings the image of distant or nearby objects into sharp focus.

To optimize the functioning of the eye, it is necessary that the amount of light that enters it via the lens be regulated. What is actually regulated is the extent of opening of the optical structure—the lens. In bright light the lens is narrowed and in dim light the lens is widened. This is effected by two antagonist muscles (circular and radial smooth muscle fibres) that form the structure of the iris diaphragm.

The phenomenological interpretation of such a system is based on the general propositions of the *theory of control.* An *input signal* that enters the system is different from that required (the so-called preset signal). The difference between the input signal and the setting is the *error signal.* The purpose of the regulating device consists in receiving the required *output signal.* The regulating system consists of a controller and an object to be controlled; it contains a feedback circuit.

Let us consider the structure of the *retina*. This is a multilayer structure shown schematically in Fig. 14.15. An image is produced on the pigmented epithelium *1*. In the outer layer *3* there is realized a synaptic coupling of photoreceptor cells *2* (rods and cones)

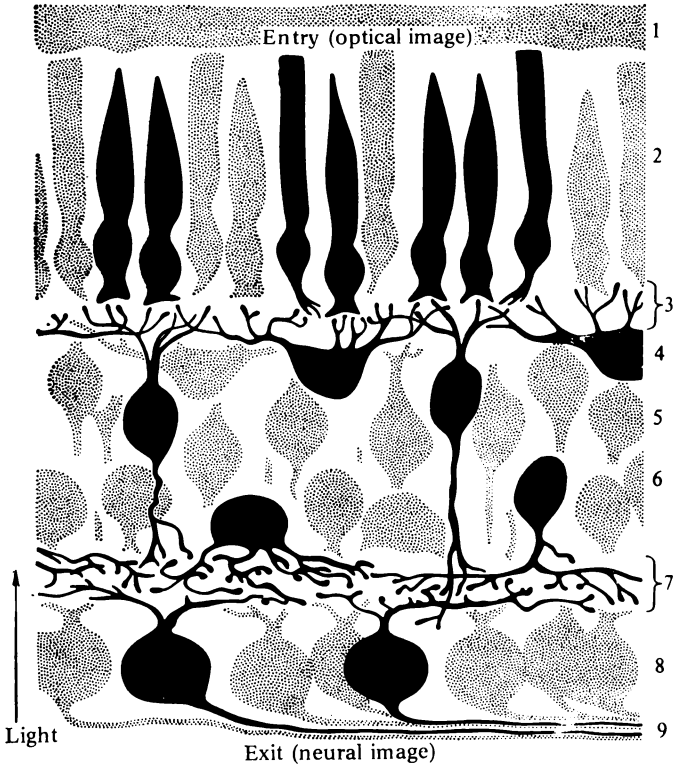


Fig. 14.15. Detailed structure of the retina:
1—outer pigmented layer (pigmented epithelium); *2*—receptor cells (rods and cones); *3*—outer synaptic layer; *4*—horizontal cells; *5*—bipolar cells; *6*—amacrine cells; *7*—inner synaptic layer; *8*—ganglion cells; *9*—optic nerve fibres.

with nerve horizontal cells *4*. The other nerve cells—bipolar cells *5* and amacrine cells *6*—are connected synaptically in layer *7* with ganglion cells *8*, which serve as direct sources of impulses that arrive at the axons of the optic nerve. The input signal is the optic image on the pigmented epithelium, and the output signal is the neural image coded by the impulses in the optic nerve. For light to reach

the photoreceptors, it must pass through the layers of nerve cells—the photoreceptor cells are thus protected against harmful effects.

The light-sensitive receptor cells, the *rods* and *cones*, are extended structures (the rods are long and narrow and the cones are short and thick) made of many specialized compartments that form a sequence of parallel disks. The rod and cone of a frog are shown schematically in Fig. 14.16. The disks contain photoreceptor molecular devices; the required substances are delivered via flagella to the disks from the basal body of the cell. Light is absorbed in the disks by the molecular system described in the next section. It has been found that the proteins of the disks are continually renewed.

It is an established fact that the cones are responsible for colour vision and the rods are sensitive to low levels of illumination. As far back as 1756 Lomonosov wrote about the visual perception of three colours—three species of ether particles—red, yellow, and blue (*The Origin of Light*, 1756). In 1802 Young proposed a theory of colour vision based on the assumption of the existence of three types of light-sensitive substances in the retina. The theory of colour vision was further developed by Maxwell and Helmholtz. This theory has been supported by direct measurements of light absorption by individual cones. There are three types of cones characterized by different curves

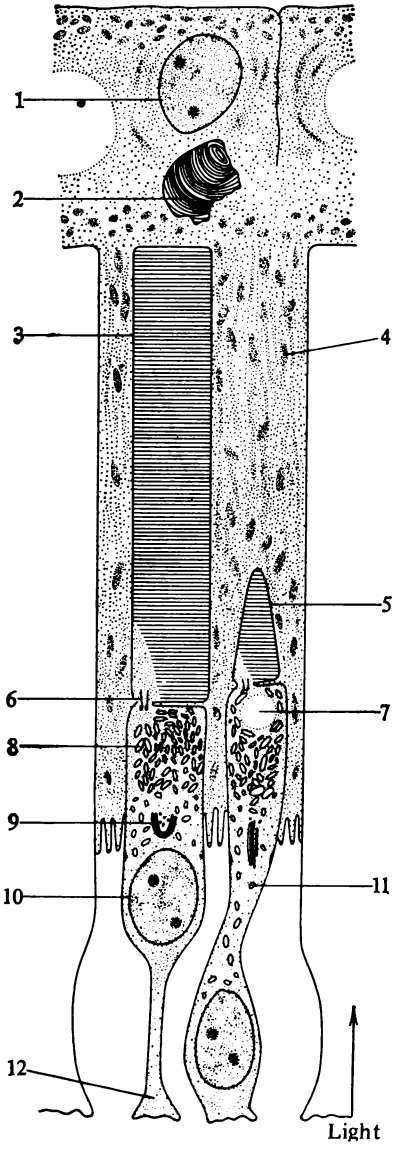


Fig. 14.16. A scheme of the structure of the rod (left) and cone (right) of a frog: *1*—nucleus of the pigmented epithelial cell; *2*—fragment peeled off from the outer segment of the rod; *3*—outer segment of the rod; *4*—pigment granules; *5*—outer segment of the cone; *6*—connector flagella; *7*—fat drop; *8*—mitochondria; *9*-Golgi complex; *10*—nucleus; *11*—ribosome; *12*—synaptic body.

of absorption of photopigments. Figure 14.17 shows these curves for primates. The curves for man are very similar to them.

The receptors of invertebrates are arranged in a different manner; in many cases they are sensitive in a wide spectral region.

The mechanisms of transformation of an optical image on the retina into a neural image, which is passed to the cerebral cortex, are rather complex. In the retina there occur the adaptation to differences in light intensities and in spectral composition of light and the perception of a three-dimensional image and of the motion

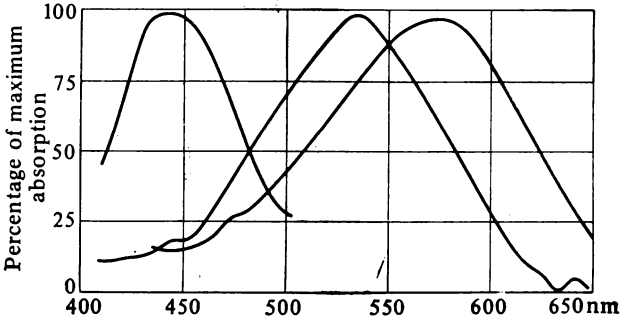


Fig. 14.17. The curves of the spectral sensitivity of the cones of primates with maxima at $\lambda = 447$ (blue-violet), 540 (green) and 577 nm (yellow).

of a visible object. There have been carried out investigations of the impulses that arise in the optic nerves of a crab and vertebrates. The crab's eye contains numerous receptors called the *ommatidia*, which resemble the rods. It was possible to study the impulses generated by individual ommatidia in the corresponding individual axons. In the dark rare periodic impulses are propagated. With light of threshold intensity additional impulses arise. If the light intensity exceeds the threshold level, then at the moment of illumination there arises a short sequence of frequent impulses. The frequency is then diminished, but it remains substantially higher than the dark frequency. When the light is switched off, a new barrage of frequent nerve impulses appear; their frequency gradually declines down to the dark frequency. In vertebrates the axons respond more strongly to changes in luminosity than during constant illumination. During strong illumination the impulses are suppressed. The retinal function is the complex, integrating nerve interaction which is reminiscent of computational work. In this sense the retina is similar to an electronic computer.

The horizontal and amacrine cells connect neighbouring photoreceptors, providing the transfer of information in the lateral direction; the bipolar cells transmit information to the inner synaptic layer. A study of the electrical activity of individual cells has shown that the receptor and horizontal cells (and in some cases, bipolar cells) undergo smooth hyperpolarization during illumination without generating a nerve impulse. In other words, their membrane potential becomes more negative. This is an unusual behaviour for a neuron. As a rule, neurons become depolarized and take on a positive membrane potential upon excitation. The nerve impulses are usually propagated in nerve cells along their length. These events do not occur in the above-indicated nerve cells of the retina. Conversely, positive nerve impulses arise in amacrine and ganglion cells. It is exactly these cells that serve as sources of impulses conducted to the cerebral cortex.

In invertebrates (a squid) the behaviour of the rods is different—they are depolarized upon illumination, i.e., they behave like ordinary excited neurons.

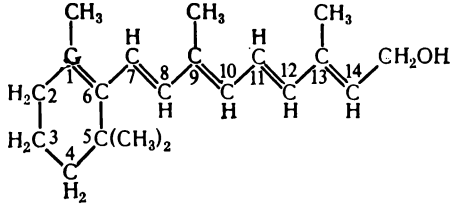
The threshold sensitivity of the eye is very high. After prolonged adaptation of the human eye to the dark it is capable of receiving individual quanta and in this respect it surpasses any photocell. Because the acts of emission of individual atoms and molecules are independent of each other with a sufficiently weak source of light the eye proves to be capable of observing *quantum fluctuations* of radiation. This was shown for the first time by Barnes and Cherny in 1932.

In 1933 Vavilov and coworkers conducted detailed visual investigations of quantum fluctuations with the aid of a very reliable procedure. They obtained precise characteristics of the rod sensitivity of the eye and important data on the nature of light. Later Hecht published his works devoted to the same problem.

14.6. The Molecular Mechanism of Photoreception

The man and higher animals perceive light of ordinary intensity in the region of about 400 to 760 nm. Ultraviolet light is absorbed by the transparent tissues of the eye. Infrared rays are not perceived by the retina. If they were perceived, an intense background of infrared radiation hindering the reception of informative signals would be produced in the warm-blooded animals. Hence, a pigment or pigments of the photoreceptor cells must absorb light in the visible region of the spectrum, i.e., must be coloured. Since we are speaking of the molecules of organic compounds, it follows that these must be molecules with a sufficiently long system of π-bonds—with a small number of conjugated bonds only ultraviolet radiation is absorbed (page 153).

Indeed, in 1933 Wald discovered the presence of *vitamin A* in the retina, in the *visual purple* or *rhodopsin.* The structure of vitamin A had been established earlier. Vitamin A (more exactly, vitamin A_1) or *retinol*$_1$ has the structure



The general formula of retinol$_1$ is $C_{19}H_{27}CH_2OH$. This molecule constitutes the almost unchanged half of the carotene molecule (see page 468). The carotenoids are a source of vitamin A, the absence of which results in blindness. The retina of vertebrates also contains vitamin A_2, or *retinol*$_2$, whose structure differs from that of vitamin A_1 by the presence of a double bond between carbon atoms 2 and 3. The general formula of retinol$_2$ is $C_{19}H_{52}CH_2OH$. The retinols are alcohols. In the photoreceptors the retinols are transformed by means of alcohol dehydrogenase (ADH) into the aldehydes—*retinal*$_1$ and *retinal*$_2$, respectively, and are linked to phospholipids and proteins. In the retinals the terminal group of the molecule is $=CH—CH=O$ and not $=CH—CH_2OH$; the general formulas of retinal$_1$ and retinal$_2$ are $C_{19}H_{27}CHO$ and $C_{19}H_{25}CHO$, respectively. The retinals contain one conjugated π-bond more than the retinols.

The retinals serve as the chromophore groups of the visual pigments containing lipoproteins called *opsins.* The opsins of the rods are different from those of the cones. Accordingly, there are formed 4 principal types of photopigments characterized in Table 14.3.

The absorption of light in a photoreceptor must be followed by a certain photochemical reaction. In *rhodopsin* the retinal forms a

Table 14.3. The Photopigments

Process		Absorption maximum, nm
$\text{Retinol}_1 \underset{\text{DPNH}}{\overset{\text{DPN}^+}{\rightleftharpoons}} \text{Retinal}_1$ (Alcohol dehydrogenase)	$+\text{Rod opsin} \overset{\text{Light}}{\rightleftharpoons} \text{Rhodopsin}$	500
	$+\text{Cone opsin} \overset{\text{Light}}{\rightleftharpoons} \text{Iodopsin}$	562
$\text{Retinol}_2 \underset{\text{DPNH}}{\overset{\text{DPN}^+}{\rightleftharpoons}} \text{Retinal}_2$ (Alcohol dehydrogenase)	$+\text{Rod opsin} \overset{\text{Light}}{\rightleftharpoons} \text{Porphyropsin}$	522
	$+\text{Cone opsin} \overset{\text{Light}}{\rightleftharpoons} \text{Cyanopsin}$	620

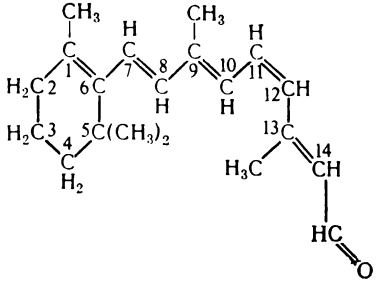
Schiff base as a result of the interaction of the aldehyde group of retinal with the amino group of opsin, namely with the ε-NH_2 group of lysyl:

$$C_{19}H_{27}CH{=}O + H_2N{-}\text{opsin} \xrightarrow{-H_2O} C_{19}H_{27}CH{=}N{-}\text{opsin}$$

The bovine rhodopsin has a molecular mass of 40,000. If we assume the molecule to be spherical, its diameter will be 4 nm. The size of the chromophore is only twice as small.

Under the action of light retinal is split off from rhodopsin; analogous processes occur also in the other three cases (see Table 14.3). This is accompanied by the photoisomerization of retinal.

An opsin-retinal complex is formed due to the CH=N bond and a number of weak interactions between the chromophore and the protein. These interactions are optimized as a result of the structural correspondence between retinal and opsin. Retinal can exist in the form of a number of isomers, since the double C=C bond forms a *trans*- or *cis*-conformation. The most stable is the all-*trans*-configuration. Such a configuration is shown for retinal on page 490. It has been found that when complexed with opsin in the dark retinal has an 11-*cis*-configuration:



In this form the CH_3 group at C_{13} and the hydrogen at C_{10} are very close to each other and a strong interaction occurs between them. As a result, the molecule is somewhat twisted and the π-bonds are brought out of the planar arrangement. The energy of conjugation is thus reduced and the 11-*cis* form must be less stable than the all-*trans* form. However, in the opsin-retinal complex the loss in conjugation energy is compensated for by a gain in energy of interaction with the protein which is greatest for the 11-*cis* form.

When the pigment is bleached in the light, retinal is split off from the opsin and isomerized to the most stable all-*trans* form. The *cis-trans* isomerization cycle is a necessary part of any photoreceptor system known.

In photoreception we deal not with a conformational change of retinal but with a *cis-trans* isomerization—rotations occur about

double bonds and not about single bonds. The difference between rotamerization and *cis-trans* isomerization is a quantitative one. In the first case, for a conformational conversion to occur, a barrier of the order of tens of kJ/mole must be overcome (12.6 kJ/mole in ethane); in the second, a barrier of about 105 kJ/mole must be surmounted for the change from 11-*cis* retinal to the all-*trans* form. The isomers of retinal are stable in solution. The differences in their energies are not large, of the order of 4-12 kJ/mole.

When the molecule is excited by the absorbed light, the π-electronic shell undergoes rearrangement and the barrier to internal rotation is lowered to a value commensurate with kT. The lifetime of the

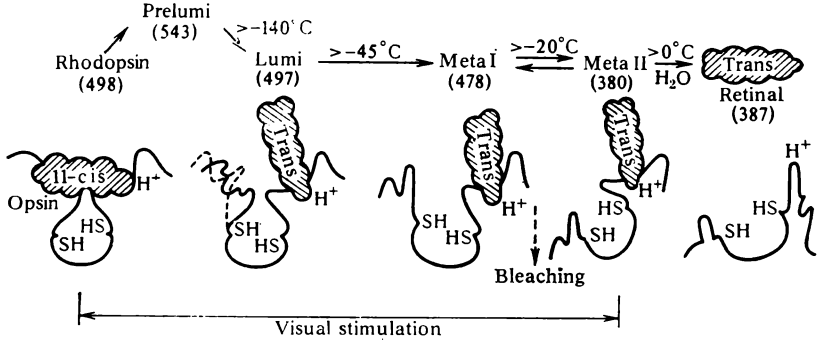


Fig. 14.18. The steps of the bleaching of rhodopsin.
The retinal is shaded. Enclosed in parenthesis are the values of λ_{max} in nm.

molecule in a singlet excited state is 10^{-9} s; the time required for rotation is 10^{-11} to 10^{-12} s. The quantum yield for the isomerization reaction 11-*cis* → all-*trans* is 0.5-0.7. The energy of a light quantum is more than sufficient for the isomerization—to a quantum with $\lambda = 500$ nm there corresponds an energy of 210 kJ/mole.

When a solution of rhodopsin in an equimolar mixture of glycerol and water is exposed at a temperature of —190°C to light of different wavelengths, the equilibrium of the isomers can be shifted reversibly, which is reflected in shifts of the absorption band. On the basis of investigations of this kind Wald constructed a scheme of successive conversions of rhodopsin shown in Fig. 14.18. Under the action of light the retinal complexed with opsin is isomerized and prelumirhodopsin is formed. These events destroy the structural correspondence between retinal and opsin and the latter together with retinal undergoes a conformational change. The structure of opsin is gradually opened, this leading to steps marked as lumirhodopsin and metarho-

dopsins. At the final stage, the Schiff base is hydrolyzed and retinal is split from opsin. When opsin is opened, new chemical groups are exposed, in particular two SH groups and the group that binds a proton with a pK value of the order of 6.6 (probably, the imidazole group).

Thus, photoreception is reduced to the isomerization of the chromophore, this being followed by a change in the structure of the lipoprotein opsin. The conformational conversion of opsin leads, in the long run, to the generation of a nerve impulse.

It has been shown that the bleaching and regeneration of rhodopsin *in vitro* depend on the frequency of short flashes of laser illumination. Only that light is absorbed whose polarization corresponds to a certain orientation of the rhodopsin molecule. It means that the ratio of the flash frequency to the time of rotation of the rhodopsin molecule in solution is essential.

Different forms of photopigments are characterized by different absorption spectra. These differences are not reduced to those indicated in Table 14.3. The absorption band of chromophore depends substantially on its interaction with opsin and, hence, on the state of the lipoprotein and its species specificities. For example, the frog retina contains two types of rods ("red" and "green") and two types of cones. The red rods contain "classical" rhodopsin, while the green rods contain rhodopsin with an absorption band shifted to the short-wavelength end. The spectral sensitivity of photopigments correlate, as a rule, wih the spectral composition of light in a medium in which a given organism exists. On going from bluish seawater to yellowish fresh water rhodopsin in the rods of fish is gradually replaced by porphyropsin. Sea fish has a sensitivity maximum of about 505; the sensitivity of freshwater fish is near 540 nm.

The facts and hypotheses briefly described here pose a number of problems that are still far from being solved. The structure of opsin and, hence, of rhodopsin, has been studied little. No detailed model of rhodopsin and of its conversions in the light can be constructed at present. The quantum chemistry of these phenomena would be of considerable interest.

It is essential that a relationship be established between the events that take place in rhodopsin and the generation of electrical potentials in the successive neurons of the retina, which results in the relay of a nerve impulse to the brain. We know very little about this relationship.

As has already been said, the photoreceptor cells of vertebrates contain stacks of lamella or disks—membranes, in which the photopigments are localized. The number of such disks in a single cell is large, amounting to 500-1000. The membranes contain an ordinary lipid bilayer. The rhodopsin molecules reside on the inner hydrophilic surface of the membrane.

The photoreceptor membranes of the outer segments of bovine and frog rods are about 60 percent protein and 40 percent lipid. Rhodopsin is the principal fraction of the protein, making up to 80 percent of its overall composition. Lipids are present mainly as phospholipids—phosphotydyl ethanolamine, phosphotydyl choline, and phosphotydyl serine. There is also 5-10 percent of glycolipid.

The disks of the outer segments of the rods are "freely floating" intracellular organelles, similar to mitochondria. The disk membrane occupies a certain space separated from the intracellular space of the photoreceptor, in which the disks are "suspended". The membrane is impermeable to Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, and PO_4^{3-}. *In vitro* the isolated disks undergo swelling and contract in response to changes in osmotic pressure of the surrounding medium.

As a result of illumination, the membrane becomes more permeable—it does not hold all the solute so effectively as does the membrane adapted to the dark. The conformational change of rhodopsin leads to a change in the state of the membrane and to an increase in passive permeability to Na^+ and K^+ ions.

X-Ray structural analysis and electron microscopy have shown that upon illumination the disk membrane undergoes structural changes, the scale of which is proportional to the time of exposure. These changes consist mainly of the translocation of rhodopsin. As a result of illumination, rhodopsin is transferred from the interdisk hydrophilic surface into the internal hydrophobic phase of the membrane. Such translocations are equivalent to phase transitions. Phase transitions have been detected in the photoreceptor cell membranes of frogs and squids and studied by means of differential calorimetry.

The photoreceptor membrane with a spin probe inserted into it has been studied by electron paramagnetic resonance. When a suspension of membranes is illuminated, there are observed significant changes in the EPR spectra, indicating an increase in the time of rotational diffusion τ of the probe in the membrane. The shape of the curve showing the dependence of τ on the duration of illumination is evidence that the photolysis of the pigment is coupled with the change of the conformational state of the membrane and with an increase in microviscosity of its hydrophobic regions. These results may be regarded as a direct proof of the conformational conversions of the membrane upon illumination.

Similar conclusions are arrived at in the study of photoreceptor membranes by gamma-resonance spectroscopy (the Mössbauer effect). The label used was iron ascorbate enriched in the isotope ^{57}Fe and the source of γ-radiation was ^{57}Co.

Of special interest is the ATPase enzymatic activity inherent in the photoreceptor membrane. As early as the thirties Engelhardt presumed that the energy of ATP could be used in the primary process of vision. This assumption was indirectly confirmed by Venk-

stern who studied the ATPase activity of the retina. It was found later that the ATPase activity is localized in the membranes of the outer segment of the photoreceptors.

The ATPase is Mg-activated, being associated with rhodopsin. The photolysis of the photopigment *in vitro* leads to noticeable inhibition of ATPase, this being presumably a consequence of an intramolecular rearrangement of the pigment which undergoes photodenaturation at the final stages of photolysis.

One might expect that the ATPase activity is necessary in vision for amplifying the signal received by the photoreceptor cell at the expense of the energy of ATP. The human eye has, as has been said above, a high sensitivity to light, being capable of recording several photons. The rod can be stimulated by a single photon with an energy of the order of 4×10^{-19} J (at $\lambda = 600$ nm). According to the Einstein basic law of photochemistry, one photon can cause a rearrangement of only one rhodopsin molecule out of 10^9 molecules contained in the rod. The light sensitivity of the rod is maximal.

The energy of one quantum is, however, insufficient for the physiological stimulation of the receptor cell. Calculations show that for this to be accomplished, amplification by several thousands of times is required. Presumably, this amplification is of the biological nature and occurs as a result of the enzymatic hydrolysis of ATP. The mechanism of this hypothetical process still remains completely obscure.

As has been recently found in the works of Skulachev, Ostrovsky, and their coworkers, the function of rhodopsin consists in that it is a *photogenerator* that produces a kind of photocurrent. Under the action of light rhodopsin gives rise to a proton concentration difference. This causes the opening of the pores in the disks and the extrusion of Ca^{2+} ions, which in turn leads to the closing of sodium channels on the outer side of the membrane and to an increase in the value of $\Delta\psi$ on it. The potential difference is responsible for the generation of a nerve impulse. The function of visual rhodopsin is found to be similar to the function of *bacteriorhodopsin* (see Sec. 14.8). In the both cases, we deal not with a *photosensor* but with a photogenerator of ΔpH. Rhodopsins are *proton pumps*, whose operation is described by the Mitchell theory.

It is exactly through the use of photoreceptor membranes that the liquid (liquid-crystalline) properties of the membranes have been best studied (cf. page 351). This has proved possible because in the photoreceptor membranes there functions only one protein, rhodopsin, which can easily be followed.

It has been established that the rods adapted to the dark exhibit dichroism. The light polarized at right angles to the long axis of the rod is absorbed several times more strongly than the light polarized in parallel to that axis. Hence, the chromophore of rhodopsin is

oriented in parallel to the plane of the disk membrane. However, on this plane the chromophores are oriented randomly and no dichroism occurs. It is possible to photoinduce dichroism by partially bleaching rhodopsin with the aid of plane-polarized light. Under ordinary conditions no dichroism appears, which is accounted for by the Brownian rotational motion of rhodopsin molecules in the retina. But when the retina is treated with glutaraldehyde, a strong photoinduced dichroism is observed. Glutaraldehyde forms cross-links which hinder the rotation of rhodopsin molecules. No energy can be transferred between rhodopsin molecules, the mean distance between which are, on an average, about 7 nm. The rotational mobility of rhodopsin in the retina has been proved. This is also evidenced by a study of transient photodichroism during pulsed photolysis. Upon pulsed illumination there occurs a jumpwise increase in absorption determined by the conversion of rhodopsin to prelumirhodopsin. This increase is much greater if the polarization vectors of the acting light and of the light whose absorption is to be studied are mutually parallel than in the case where these vectors are mutually perpendicular. A flash induces dichroism, which then quickly disappears. The half-time of disappearance of dichroism at 20°C is 3.0 ± 1.5 μs. As the temperature increases this process is accelerated. An increase in the viscosity of the medium retards the disappearance of photodichroism.

If the dependence on time is really determined by rotational diffusion, then it is described by the following equation:

$$\frac{\partial^2 n}{\partial \vartheta^2} = \tau \frac{\partial n}{\partial t} \tag{14.7}$$

where n is the fraction of chromophores oriented in the interval from ϑ to $\vartheta + d\vartheta$; ϑ is the angle made by the chromophore with the electrical vector of the flash light; τ is the relaxation time. For a linear chromophore and rotational diffusion about the axis perpendicular to the disk membrane, solution of Eq. (14.7) has the form

$$n = 1 + fe^{-4t/\tau} \cos 2\vartheta \tag{14.8}$$

The time t is measured from the onset of a flash; $f \leqslant 1$ is an empirical factor. The dichroic ratio is equal to

$$\frac{\int_0^{\pi/2} n(\vartheta) \cos^2 \vartheta \, d\vartheta}{\int_0^{\pi/2} n(\vartheta) \sin^2 \vartheta \, d\vartheta} = \frac{2 + fe^{-4t/\tau}}{2 - fe^{-4t/\tau}} \tag{14.9}$$

Under ideal conditions when $f = 1$, the maximum initial ratio is equal to 3. This corresponds to the value found for the retina treated with glutaraldehyde. In pulsed experiments there was observed a value of 2 corresponding to $f = 0.7$. The time of rotational diffusion

τ depends on the viscosity of the medium η, temperature, and the size of the chromophore molecule. According to Einstein, for a sphere of radius r, which experiences Brownian rotation about a certain axis,

$$\tau = \frac{8\pi r^3}{kT}\eta \qquad (14.10)$$

At 20°C the time τ for rhodopsin has a value of about 20 μs and r = 2.2-2.8 nm; rhodopsin is nearly spherical and is embedded in the membrane. It follows that the viscosity of the membrane η is about 0.2 Pa·s (the interval is from 0.07 to 0.6 Pa·s), i.e., it is close to the viscosity of a light oil, say, olive oil.

During the rotations of the rhodopsin molecule, which arise because of its conformational conversions, the degree of embedding of rhodopsin in the liquid bilipid membrane is changed, which is essential for the change of its ionic permeability.

The diffusion constant found for the rods from the frog retina is equal to $3.5 \pm 1.5 \times 10^{-9}$ cm^2·s^{-1}. The viscosity of the membrane is found to lie within the same interval of values, between 0.1 and 0.4 Pa·s. The time elapsed between the collisions of neighbouring rhodopsin molecules in the retina, τ_{col}, is found from the formula

$$s^2 = 4D\tau_{col} \qquad (14.11)$$

where s is the distance between the rhodopsin molecules. If the effective diameter of rhodopsin is 4.5 nm and the distance between the centres of the molecules is 7 nm, then s = 2.5 nm and τ_{col} = 4 μs, which is 5 times less than the relaxation time of rotational diffusion. The collision frequency is 10^5 to 10^6 s^{-1}.

These results, obtained by Cone *et al.*, are very interesting. They enable one to construct the theory of functioning of the membranes based on their liquid properties (cf. page 351). We still know little about the events leading to the generation of a nerve impulse in the photoreceptor system. These events may be expected to be associated with the behaviour of the liquid membrane.

14.7. Bacteriorhodopsin

Rhodopsin functions not only in the eye retina. A certain region of the membrane surface of red halobacteria (*Halobacterium halobium*) contains the protein bacteriorhodopsin (BR). This protein has a molecular mass of 26,000 and is composed of one protein molecule attached to a retinal molecule via an aldimine bond. The absorption maximum of bacteriorhodopsin is at 570 nm. *Halobacteria* are extremal halophiles which can grow in saline lakes. They have a positive phototaxis to visible light and a negative phototaxis to ultraviolet light.

Together with lipids bacteriorhodopsin constitutes a system that provides the transport of protons across the membrane against the pH gradient. Bacteriorhodopsin functions as a photoinduced proton pump (Oesterhelt and Stockenius, 1971).

The operation of the proton pump in halophiles is coupled with the phosphorylation of ATP and with the operation of the sodium pump, which provides the desalting of the internal medium of the bacterial cell. The Mitchell theory (Sec. 13.3) could possibly be applied to this system. As pointed out earlier (page 495), rhodopsin also functions as a photogenerator, performing the role of the proton pump.

In contrast to the membranes of the rods and cones, the purple membrane of bacteria is crystalline rather than liquid. The molecules of bacteriorhodopsin are assembled into clusters (three molecules in a cluster) and packed into a hexagonal lattice, which resides in a compact lipid matrix. In the middle of the cluster there is a channel 1-2 nm in diameter. The bacteriorhodopsin molecule is 70-80 percent α-helical. Each molecule contains 7 rod-like α-helices 3.5-4 nm long, which are oriented perpendicular to the membrane plane. The size of the protein is $2.5 \times 3.5 \times 4.5$ nm^3. Bacteriorhodopsin runs through the entire membrane and comes into contact with aqueous phases on both sides of the membrane. Bacteriorhodopsin molecules are asymmetrical (see page 349). From the value of the transition dipole moment of retinal (12 D) it follows that on light absorption an electron is transferred to a distance comparable with the distance between the terminal groups of retinal—the cyclohexene ring and the Schiff base.

The photoinduced proton transport involves changes in the conformation of bacteriorhodopsin and its spectrum. The light-adapted form BR_{570} enters the cycle of photoinduced reactions, giving rise to a number of intermediate compounds, this being accompanied by the transport of H^+ across the membrane and by the regeneration of the initial form BR_{570}. The cycle is schematically shown in Fig. 14.19. As a result of one revolution of the cycle, upon absorption of one light quantum one proton is passed through the membrane. In a certain phase of the cycle BR releases a proton into the acidic medium (pH = 3.0) on one side of the membrane. In another phase, the same group absorbs the proton from the alkaline medium (pH = = 12.0) on the other side of the membrane.

Chernavskaya and Chernavsky (1978) examined this process on the basis of the tunnelling transport of electrons leading to an intramolecular transformation of energy (see Sec. 13.4). The model consists of two potential wells separated by a barrier (Fig. 14.20). The wells correspond to electrophilic groups—the cyclohexene ring of retinal and the nitrogen of lysyl (attached to retinal in a Schiff base). In well *I* there are two levels, the distance between which is

equal to the energy of the quantum absorbed. The level of the second well is somewhat lower than the upper level of the first well. The electron is excited by light in well *I*, tunnels from the upper level into well *II*. As a result of the approach of the proton, well *II* becomes polarized, this being followed by a lowering of the electronic level. Being positively charged, well *I* is also polarized—the electronic levels in it are raised. A further lowering of the level in well *II* and a rise in well *I* occur as a result of conformational changes resulting in the relaxation to a new equilibrium state corresponding

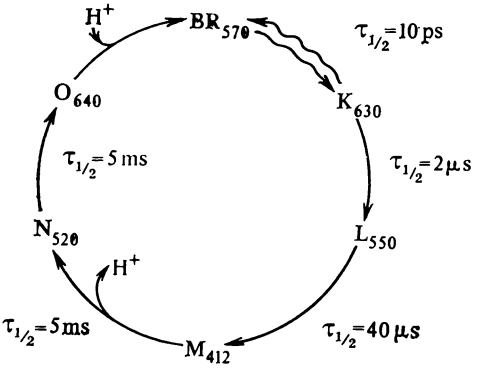


Fig. 14.19. The functioning cycle of bacteriorhodopsin.

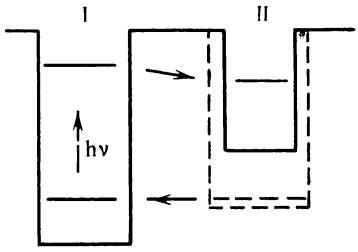


Fig. 14.20. Two electrophilic groups separated by an energy barrier.

to a new charge distribution. This is followed by the reverse tunnelling of the electron from well *II* to the lower level of well *I*, by a repeated polarization of both wells and a conformational relaxation to the initial state. The cycle is thus terminated. The resonance conditions required for tunnelling are provided by polarization and conformational transitions.

The free-energy change of the proton upon tunnelling of an electron from *I* to *II* is given by

$$\Delta G = \frac{e^2}{\varepsilon}\left(\frac{1}{r_2} - \frac{1}{r_1}\right) \qquad (14.12)$$

where $e = 4.8 \times 10^{-10}$ esu is the electronic charge; $\varepsilon \approx 5$ is the dielectric constant; r_2 is the distance from the proton to the centre of well *II*; r_1 is the distance to the centre of well *I*. For a proton that resides near the nitrogen atom of the Schiff base and participates in its protonation $r_2 = 0.25$ nm and $r_1 = 1.3$ nm; so $\Delta G \approx 1$ eV.

The corresponding change in the dissociation constant of this group is $\Delta pK = \Delta G/2.3\ kT = 17$. The change of pK on reverse tunnelling is of the same order of magnitude but of opposite sign. These estimates show that the alteration of the properties of the group under consideration brought about by electron tunnelling can occur over the entire interval of the pH values observed—from 3 to 12.

The lowering of the electronic level for polarization is expressed in a similar manner:

$$\Delta E = \frac{e^2}{\varepsilon}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{14.13}$$

where r_1 and r_2 are the distances from the proton to the centre of the well before and after polarization. At $r_2 = 0.25$ nm and $r_1 = 0.15$ nm we find that $\Delta E = 0.75$ eV. The shifts of the levels caused by conformational changes are of the same order. The approach of the levels required for the reverse process can be accomplished.

The process involves the protonation of the Schiff base, the expulsion and capture of this and other protons, the polarization and rotamerization of retinal and rhodopsin. These phenomena cannot be considered to have been thoroughly studied, despite the number of interesting spectroscopic and kinetic investigations that have been carried out. The scheme proposed by Chernavskaya and Chernavsky does not contradict the available results but direct supporting evidence is required.

Chapter 15 Modelling of Dynamic Biological Processes

15.1. Dynamic Order

In the preceding chapters we have repeatedly dealt with the nonlinear behaviour of a biologically functional system. It suffices to recall the generation and propagation of a nervous impulse (Chapter 11). When the threshold of stimulus strength is reached, the system is switched over to a new state—a nerve impulse is started. The initiation of a nerve impulse is comparable to pulling the trigger of a gun or lighting a fuse. Similar trigger phenomena occur in all cases of the emergence of an ordered behaviour of biological systems in space and time.

We have already mentioned the differences between the equilibrium, static order and the dynamic order, the latter being inherent in far-from-equilibrium open systems called *dissipative systems* (Sec. 9.7). Here we shall deal once again with this central topic of theoretical biophysics. In this and subsequent chapters we shall be concerned with a number of models of biological dissipative systems on the basis of general theoretical approaches to their behaviour. In such systems there arise processes of self-organization in space and time. We have pointed out earlier that the branch of natural sciences that deals with such processes is known as *synergetics*.

We shall begin with the thermodynamic characteristics of three types of systems commonly encountered in biophysics.

The behaviour of an isolated system is fully characterized by the second law of thermodynamics in its canonical form—the entropy of the system tends to a maximum. Neither self-organization nor phase transitions are possible in such a system.

A closed system, which exchanges energy but not matter with the outside world, is capable of undergoing phase transitions to a static, equilibrium, ordered state. The system is characterized by a free energy $G = H - TS$ which tends to a minimum. At a sufficiently low temperature, the entropy contribution to the free energy becomes small and there arises, for example, a static crystalline order.

In the case of open systems, one should clearly distinguish two types of behaviour. The behaviour of a near-equilibrium system is

described within the framework of linear thermodynamics (Chapter 9). Here we deal with equilibrium structures modified as a result of restrictions that hinder the attainment of equilibrium. The appearance of a dynamic order in these conditions is impossible—the steady states are asymptotically stable, which is illustrated by Fig. 15.1.

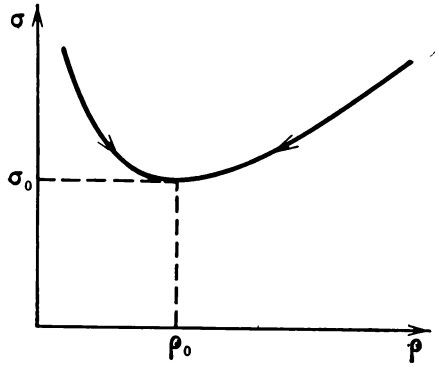


Fig. 15.1. The dependence of σ on ρ near equilibrium.

The dependence of the dissipation function σ on the parameter ρ that defines the system has a minimum value σ_0 in the stationary state $\rho = \rho_0$ (the Prigogine theorem, page 330). When ρ deviates from ρ_0, the system returns to the state ρ_0 exponentially, without experiencing oscillations.

Finally, in a dissipative system, i.e., an open far-from-equilibrium system, there arises a dynamic order, a coherent behaviour of the ensemble when the values of the system's parameters corresponding to *instabilities* are exceeded.

The situations typical for open far-from-equilibrium systems are also described in terms of the macroscopic thermodynamic variables. A dynamic order appears as a result of the increase of fluctuations up to the macroscopic level. There is a far-reaching and rather informative analogy between these processes and *phase transitions*—the formation of a new biological structure, a new species, is a kind of phase transition (see Sec. 15.5).

The criterion for the possibility of the emergence of dynamic order in a dissipative system is the failure to satisfy stability conditions. Let us cite once again some of the relations of the thermodynamics of open systems. The *dissipation function* is given by

$$\sigma = \frac{d_i S}{dt} = \sum_j J_j X_j \geqslant 0 \qquad (15.1)$$

Its change in time is written in the form

$$\frac{d\sigma}{dt} = \frac{d_J\sigma}{dt} + \frac{d_X\sigma}{dt} = \sum_j X_j \frac{dJ_j}{dt} + \sum_j J_j \frac{dX_j}{dt} \tag{15.2}$$

Under constant boundary conditions we have

$$\frac{d_X\sigma}{dt} \leqslant 0 \tag{15.3}$$

This yields the stability condition for the steady state under consideration:

$$\sum_j \delta J_j \delta X_j \geqslant 0 \tag{15.4}$$

where δJ_j and δX_j are the deviations of generalized fluxes and forces from their steady-state values. Near equilibrium this condition is fulfilled at all times. As applied to chemical processes it has the following form:

$$\sum_j \delta v_j \delta \mathcal{A}_j \geqslant 0 \tag{15.5}$$

If condition (15.4) is not satisfied, the steady state is unstable and the fluctuations may be amplified, leading to the appearance of dynamic order. "Order through fluctuations" is evidently possible only in such an open system whose behaviour is essentially nonlinear.

In biology we deal with several types of phenomena which serve as a direct evidence of the nonlinearity of the corresponding processes.

First, there are all-or-none threshold processes involving the switchover of the system from one regime to another, say, the generation of a nerve impulse or muscle contraction.

Second, the behaviour of cells and organisms at all levels of organization is subject to regulation and control, which are determined, say, by feedback mechanisms which are absent in linear systems.

Third, there are periodic, oscillatory phenomena. At all levels of organization, from the macromolecular to the population level, in biological systems there occur undamped oscillations of the characteristic parameters—enzyme activity, metabolite concentration, population number.

Fourth, any biological system, from the cell to the biosphere as a whole, irreversibly undergoes evolutionary development. The development invariably implies the emergence of new structures, the creation of new information, i.e., basically nonlinear processes.

It is necessary to consider the multiple stable and unstable steady states of dissipative systems and transitions between them. As a matter of fact, thermodynamics as such leaves the stage here—the description of nonlinear systems on its basis is insufficient. Ther-

modynamics only permits the formulation of stability criteria. In order to further consider self-organizing systems it is necessary to resort to physico-mathematical modelling, to the construction of dynamic models.

Such modelling is accomplished with the aid of the apparatus of differential equations. The question of the applicability of this apparatus to biological systems is not trivial.

A "chemical machine" is, generally speaking, characterized by a discrete rather than a continuous set of states. The application of the apparatus of differential equations to such a system implies the inclusion of discrete states in a certain continuous set. This procedure does not interfere with the treatment of the behaviour of a discrete system; on the contrary, with an appropriate choice of the model it allows one to analyze it. At the same time, the apparatus of deterministic, differential continual equations may prove insufficient for the study of processes that involve a small number of molecules or species. Such processes are *stochastic*, probabilistic processes; their analysis requires the use of probability theory and, in a number of cases, of the *Markov chain theory*. The question of the mathematical apparatus must be solved separately for each class of models. Modelling itself is determined by the process under study and depends directly on the time scale in which it is progressing. Any biological system undergoes numerous nonlinear kinetic processes characterized by their intrinsic times.

In this and subsequent chapters we shall be concerned with the various models of biological processes. The principal method is the study of differential equations that describe the dynamics of the model, but in a number of cases such a study must be supplemented by the solution of the corresponding stochastic problems.

15.2. Physico-Mathematical Foundations of the Dynamics of Nonlinear Processes

The method of investigating a biological system is defined here as the construction of a dynamic model and its description by means of differential equations, i.e., the construction of a system of equations and the study of their solutions.

A sufficiently general form of a mathematical model is

$$\begin{aligned} \frac{dx_1}{dt} &= \mathcal{F}_1(x_1, \ldots, x_N) \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ \frac{dx_N}{dt} &= \mathcal{F}_N(x_1, \ldots, x_N) \end{aligned} \tag{15.6}$$

where $x_1, \ldots, x_N$ are the physical variables which define the system and which depend on time and initial conditions; $\mathcal{F}_1, \ldots, \mathcal{F}_N$ are the nonlinear functions of these variables in a general case. For example, the quantities x_i can express the concentrations of metabolites or the number of individuals.

The ordinary method of investigation of nonlinear equations of the type (15.6) consists of their linearization. First, steady-state values are sought for the variables $x_1^0, x_2^0, \ldots, x_N^0$, which are the solutions of equations (15.6) at $\dot{x}_1 = \ldots = \dot{x}_N = 0$. Then, one investigates the linear equations written in terms of the variables representing small deviations from the steady-state values $x_i - x_i^0$. Here the terms of the second order of smallness which are nonlinear with respect to $x_i - x_i^0$ may be neglected.

We begin with the study of a simple linear model—an oscillator with friction. The equation of motion has the form

$$m\ddot{x} + b\dot{x} + kx = 0 \tag{15.7}$$

It contains only one variable—the deviation from the equilibrium position x. But we may pass over to a system of two equations of the type (15.6), introducing a second variable—the velocity $y = \dot{x}$. We then have

$$\dot{x} = y, \quad \dot{y} = -\frac{b}{m} y - \frac{k}{m} x \tag{15.8}$$

Both equations are linear, like the basic equation (15.7). Its solution is written as follows:

$$x = A_1 \exp(\lambda_1 t) + A_2 \exp(\lambda_2 t) \tag{15.9}$$

where λ_1 and λ_2 are the roots of the quadratic equation (see page 510):

$$\lambda^2 + \frac{b}{m} \lambda + \frac{k}{m} = 0 \tag{15.10}$$

At $b^2 > 4km$ these roots are real; at $b^2 < 4km$ they are complex. In the first case the process has the character of aperiodic damping and in the second, that of damped oscillations. The values of A_1 and A_2 are determined by the initial conditions.

The rational method of investigation of a dynamic system consists in obtaining its "phase pattern". The behaviour of the system is visualized as the movement of a representative point in the x, y phase plane where $y = \dot{x}$. If the number of variables is greater than two, not a phase plane but a phase space must be resorted to. In many cases, however, it proves possible to consider only second-order systems of the type (15.8), i.e., $N = 2$.

In a phase plane the point moves along a phase trajectory with a phase velocity. The phase-trajectory equation for an oscillator with friction is deduced by eliminating time from Eqs. (15.8). To do this, we divide the second equation by the first:

$$\frac{dy}{dx} = -\frac{2hy + \omega_0^2 x}{y} \tag{15.11}$$

where $2h = b/m$ and $\omega_0^2 = k/m$. This equation describes integral curves, at each point of which the tangent line has a slope equal to dy/dx. Together with Eqs. (15.8), Eq. (15.11) defines on the phase

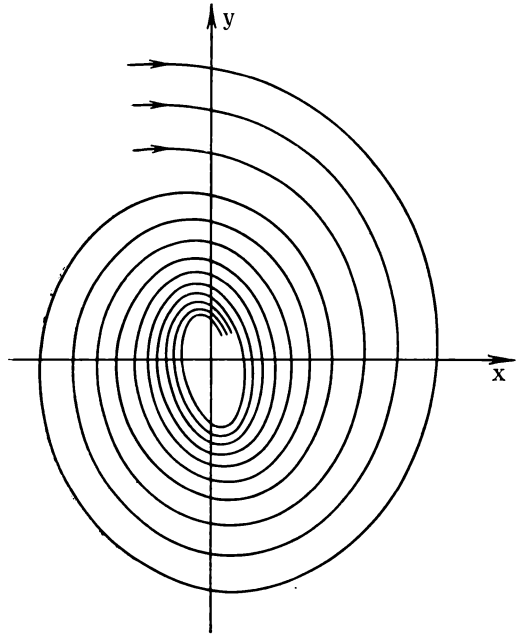


Fig. 15.2. The integral curves on the phase plane for attenuating vibrations of an oscillator with friction.

plane a certain vector field with the only singular point $x = 0$, $y = 0$. It is convenient to investigate this field with the aid of *isoclines*, i.e., curves (straight lines in this particular case) which are the geometrical loci of points at which the tangents to all the integral curves have the same slope. In the case of an oscillator the equa-

tion of the isocline with slope $\varkappa$ has the form

$$dy/dx = \varkappa$$

or

$$y = ax = -\frac{\omega_0^2}{\varkappa + 2h}\dot{x} \tag{15.12}$$

Thus, the isoclines are straight lines passing through the coordinate origin—through the singular point $x = 0$, $y = 0$.

The solution of Eq. (15.11) at $b^2 < 4k/m$ or $h^2 < \omega_0^2$ has the form of a coordinate equation:

$$y^2 + 2hxy + \omega_0^2 x^2 = C \exp\left(2\frac{h}{\omega_0} \arctan \frac{y + hx}{\omega_0 x}\right) \tag{15.13}$$

where C is a constant determined by the initial conditions. This solution may be represented by a family of logarithmic spirals shown in Fig. 15.2. The vector field constructed with the aid of isoclines is depicted in Fig. 15.3.

The phase velocity is found from the following equation:

$$\boldsymbol{v} = \boldsymbol{i}\dot{x} + \boldsymbol{j}\dot{y} \tag{15.14}$$

where $\boldsymbol{i}$ and $\boldsymbol{j}$ are unit vectors. In our case, from Eqs. (15.8) we obtain:

$$\boldsymbol{v} = \boldsymbol{i}y + \boldsymbol{j}(-2hy - \omega_0^2 x)$$

and

$$|\boldsymbol{v}|^2 = \dot{x}^2 + \dot{y}^2 = \omega_0^4 x^2 + 4h\omega_0^2 xy + (1 + 4h^2)\, y^2$$

The phase velocity decreases upon approach to the coordinate origin and vanishes at this point.

At $h = 0$, i.e., $b = 0$, there is no friction and the system becomes an undamped harmonic oscillator. The integral curves represent a family of ellipses:

$$y^2 + \omega_0^2 x^2 = \text{const} \tag{15.15}$$

The isocline equation has the form $y = -\omega_0^2 x/\varkappa$, and the phase velocity is given by $\boldsymbol{v} = \boldsymbol{i}y - \boldsymbol{j}\omega_0^2 x$.

For a damped aperiodic process $b^2 > 4k/m$, i.e., $h^2 > \omega_0^2$. The roots of the characteristic equation are

$$\lambda_{1,2} = -h \pm \sqrt{h^2 - \omega_0^2}$$

The phase pattern of the system is shown in Fig. 15.4.

In the examples described above we deal with various types of singular points lying at the coordinate origin in all the three cases. For a harmonic oscillator without friction all the phase curves are closed and have the shape of an ellipse. They surround a singular

point called a *centre*. For damped oscillations the singular point is an asymptotic point of all the curves that have the form of spirals inserted one into another. Such a point is called a *focus*. Finally,

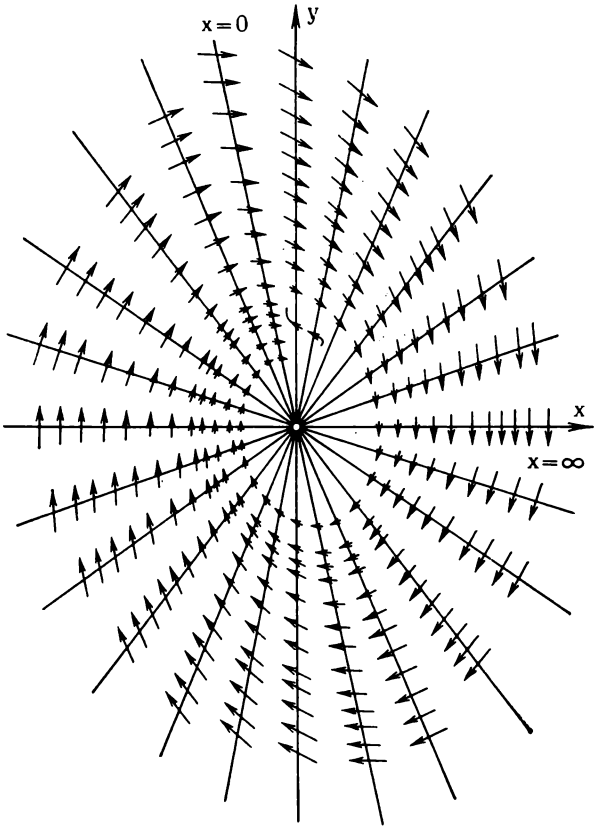


Fig. 15.3. The vector field for an oscillator with friction.

upon aperiodic damping all the curves pass through a singular point called a *node*.

Let us carry out a general investigation of singular points for a second-order system. We have two nonlinear differential equations:

$$\begin{aligned} \dot{x} &= a_1 x + a_2 y + X_2(x, y) \\ \dot{y} &= b_1 x + b_2 y + Y_2(x, y) \end{aligned} \tag{15.16}$$

where X_2 and Y_2 are polynomials containing terms of higher orders than the first in x and y. The right-hand sides of the equations vanish at the origin of the coordinates $x = 0$, $y = 0$. Hence, this is the

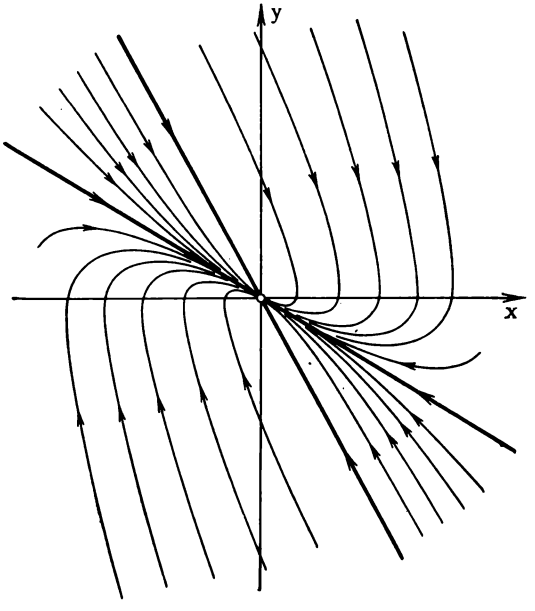


Fig. 15.4. The integral curves for the aperiodic damping of an oscillator with friction.

singular point corresponding to the steady state $\dot{x} = 0$, $\dot{y} = 0$. Restricting ourselves to a linear approximation, i.e., considering only the neighbourhood of this point, we have

$$\dot{x} = a_1 x + a_2 y \quad \text{and} \quad \dot{y} = b_1 x + b_2 y \tag{15.17}$$

The integral curve is

$$\frac{dy}{dx} = \frac{b_1 x + b_2 y}{a_1 x + a_2 y} \tag{15.18}$$

We seek solutions of Eqs. (15.17) in the form

$$x = A \exp(\lambda t) \quad \text{and} \quad y = B \exp(\lambda t)$$

Substituting these solutions into Eqs. (15.17), we obtain two homogeneous equations for x and y. The condition for their compatibility has the form

$$\begin{vmatrix} a_1 - \lambda & a_2 \\ b_1 & b_2 - \lambda \end{vmatrix} = 0$$

or

$$\lambda^2 - (a_1 + b_2)\lambda + a_1 b_2 - a_2 b_1 = 0 \qquad (15.19)$$

Since this quadratic equation has two roots, λ_1 and λ_2, and the system of equations (15.17) is linear, its complete solution takes the form

$$x = A_1 \exp(\lambda_1 t) + A_2 \exp(\lambda_2 t)$$
$$y = B_1 \exp(\lambda_1 t) + B_2 \exp(\lambda_2 t)$$

The values of A_1, A_2, B_1, and B_2 are determined by the initial conditions, i.e., by the quantities x, y, $\dot{x}$, and $\dot{y}$ at $t = 0$.

In the case of an oscillator with friction, $a_1 = 0$, $a_2 = 1$, $b_1 = -\omega_0^2$, and $b_2 = -2h$.

The general classification of singular points given by Poincaré is based on the behaviour of integral curves in the nearest neighbourhood of these points.

If the discriminant of the characteristic equation is $D = -(a_1 - b_2)^2 - 4a_2 b_1 \leqslant 0$, then both roots, λ_1 and λ_2, are real. If here $a_1 b_2 - a_2 b_1 > 0$, their signs are identical. We have the following cases:

1. $\lambda_1, \lambda_2 < 0$. The solution has the form of exponentials that decrease with time, i.e., the system withdrawn from this point returns to it. The singular point is a *stable node.*

2. $\lambda_1, \lambda_2 > 0$. The system recedes from a singular point which is an *unstable node.*

3. If $D \leqslant 0$ and $a_1 b_2 - a_2 b_1 < 0$, the roots λ_1 and λ_2 have opposite signs. The singular point is unstable and is called a *saddle.* Two integral curves pass through it; these curves are called the *separatrices.* The remaining phase trajectories go to infinity, bypassing the singular point.

4. If $D < 0$ but $a_1 b_2 - a_2 b_1 = 0$, i.e., $D = -(a_1 + b_2)^2$, then $\lambda_1 = 0$ and $\lambda_2 = a_1 + b_2$. One of the roots is equal to zero. For the linear system (15.17), instead of a singular point, we obtain a straight line corresponding to steady states at which the remaining integral curves terminate; the direction of motion along these curves depends on the sign of λ_2.

If the discriminant $D > 0$, then the roots λ_1 and λ_2 are complex-conjugate.

5. The real parts of λ_1 and λ_2 are negative, i.e., $a_1 + b_2 < 0$. Damped oscillations occur in the system; the singular point about which the spiral phase trajectories wind up is a *stable focus.*

6. The real parts of λ_1 and λ_2 are positive, i.e., $a_1 + b_2 > 0$. The singular point is an *unstable focus* corresponding to oscillations that increase in amplitude.

7. The roots $\lambda_1 = -\lambda_2$ are imaginary, i.e., $a_1 + b_2 = 0$. Undamped oscillations take place in the system and the singular point is a *centre.* The phase trajectories are concentric ellipses.

The types of singular points are shown in Fig. 15.5.

A singular point represents a stable or unstable equilibrium or steady state. The definition of the stability of the equilibrium state

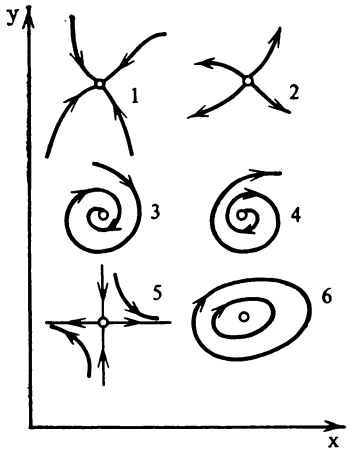


Fig. 15.5. Types of singular points: *1*—stable node; *2*—unstable node; *3*—stable focus; *4*—unstable focus; *5*—saddle; *6*—centre.

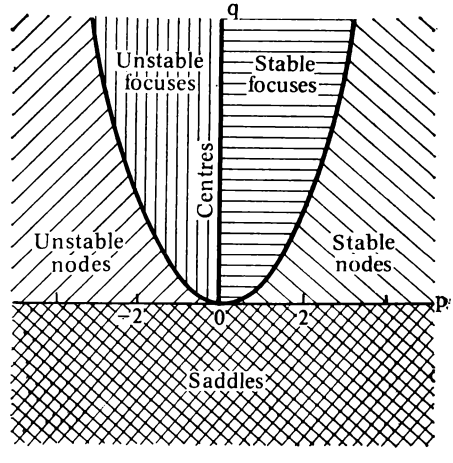


Fig. 15.6. The regions of singular points.

(which also applies to the steady state) according to Lyapounov is as follows:

The equilibrium state is stable if for any given region ε of permissible deviations from the equilibrium state there is a region $\delta(\varepsilon)$ that surrounds this state and has such a property that no motion starting inside δ reaches the boundary of the region ε. And, conversely, the equilibrium state is unstable if there is a region ε for which the region $\delta(\varepsilon)$ does not exist. Let the region ε be a square in the phase plane; then the state $x = x_0$, $y = y_0$ is stable if, by spec-

ifying in advance a positive value of ε as small as we please, we can find such $\delta(\varepsilon)$ that if at $t = 0$

$$|x(0) - x_0| < \delta \quad \text{and} \quad |y(0) - y_0| < \delta$$

then at $t \to \infty$

$$|x(t) - x_0| < \varepsilon \quad \text{and} \quad |y(t) - y_0| < \varepsilon$$

It is exactly this type of stability and instability that we deal with here.

The theory of nonlinear dynamical systems considered here has been developed by Mandelstam and Andronov.

The theory of nonlinear dynamical systems makes a distinction between structurally stable and unstable systems. In the first case, small changes in the system's parameters do not alter its general behaviour—the mathematical model is stable with respect to small changes in the form of the differential equations. The situations corresponding to cases 1, 2, 3, 5, and 6 of the above classification characterize structurally stable systems. Conversely, in cases 4 and 7 the systems are unstable. Indeed, in case 4 the value of the parameter $a_1b_2 - a_2b_1 = 0$ is critical and in going from a positive to a negative value of the parameter there appears a saddle instead of a stable node. In case 7 the value of $a_1 + b_2 = 0$ is critical: upon the transition $a_1 + b_2 > 0 \to a_1 + b_2 = 0 \to a_1 + b_2 < 0$ the singular points are changed thus: unstable focus → centre → stable focus.

The values of the parameters of the system at which its behaviour is changed are called critical or *bifurcation points.*

Let us return to the characteristic equation of the system with two degrees of freedom, (15.19). We rewrite it in the form

$$\lambda^2 + p\lambda + q = 0 \qquad (15.19a)$$

The coefficients p and q are functions of the system's parameters—the parameters a_1, a_2, b_1, and b_2 in the case under consideration. The regions of various singular points can conveniently be depicted in the p, q-plane (Fig. 15.6). The roots λ_1 and λ_2 have a negative real part only at $p > 0$, $q > 0$. The complex roots that correspond to the foci are in the region of $q > p^2/4$ only, i.e., between the branches of the parabola $q = p^2/4$, and the region of $q < p^2/4$ corresponds to the nodes. The centres are on the positive side of the ordinate—at $p = 0$, $q > 0$. When the system's parameters are changed, the representative point can cross the boundary of the region. In this case, bifurcation occurs.

Special cases of the critical points that correspond to various types of structural instabilities are dealt with in René Thom's *theory of catastrophes.* This theory is associated with the theory just described. It is so far difficult to indicate the branch of biophysics in which valuable results have been obtained with the aid of the theory of catastrophes.

Among the linear systems considered so far, only the harmonic oscillator without friction has closed phase trajectories corresponding to a periodic motion. In such a system the energy is constant, the system being conservative. No periodic processes can occur in linear nonconservative systems.

The oscillatory behaviour of nonlinear systems is very complicated and varied. Its study is of fundamental value for a very wide range of physical problems, including the problems of biophysics.

The general equations (15.6) are nonlinear; the same is true for equations (15.16) which we linearized with the purpose of investigating the neighbourhood of singular points. Such an investigation,

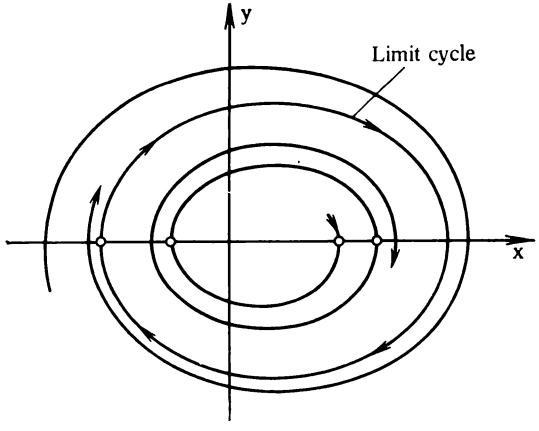


Fig. 15.7. The limit cycle.

however, fails to provide answers to the questions of the behaviour of a nonlinear system over the entire phase plane. In the subsequent discussion we shall encounter nonlinear systems defined by multiple singular points.

Of special interest for biology are nonlinear *auto-oscillatory systems* in which undamped oscillations are set up and maintained, despite the presence of friction. This occurs due to the forces dependent on the state of motion of the system itself. The amplitude of auto-oscillations is determined by the properties of the system and not by initial conditions. From the unstable singular points the phase trajectories go to infinity or to stable points. But in the case of auto-oscillations these trajectories wind up around the closed curve surrounding the singular point—around the *limit cycle* (Fig. 15.7). The limit cycles in turn may also be unstable. These

situations are inherent in characteristic for structurally stable systems, to which presumably belong a number of biological systems. The stable limit cycle depicted in Fig. 15.7 corresponds to undamped auto-oscillations. We have already dealt with such a phenomenon in discussing the properties of the flight muscles of insects.

The special features of chemical and biological dynamical systems of interest are as follows.

1. The dynamical variables in chemistry and in a number of biological problems are the concentrations of the reactants. On the same basis, the numbers of organisms (in population genetics and ecology) are also treated as variables.

2. In a chemical-biological system the chemical processes are associated with diffusion processes, with transport of matter. In other words, here we deal not with point systems but with *distributed systems*. Of special value for biology is *compartmentation*—the division of the systems into compartments separated by membranes. The system is heterogeneous not only chemically but spatially as well.

3. In living chemical systems the nonlinear chemical reactions are coupled with the transport of matter and also with mechanical and electrical processes.

4. In many chemical-biological processes we have to deal with small numbers of molecules. The concentration concept itself has limited applicability in these cases, and the probabilities of the various states of molecules should be introduced as dynamical variables.

The mathematical models of biological processes are often very complicated, contain many variables and describe a multistage behaviour. There are, however, possibilities for a simplified treatment which consists in reducing a large set of equations to two or three nonlinear equations.

Suppose that with a number of transformations performed and appropriate scales chosen we have succeeded in arranging the system of nonlinear equations (15.6) in powers of the small parameter $\varepsilon \ll 1$ at the derivative

$$\begin{aligned} \varepsilon^2 \dot{x}_i &= \mathcal{F}_i(x_1, \ldots, x_N), \quad i = 1, 2, \ldots, l \\ \varepsilon \dot{x}_j &= \mathcal{F}_j(x_1, \ldots, x_N), \quad j = l+1, \ldots, l+m \\ \dot{x}_k &= \mathcal{F}_k(x_1, \ldots, x_N), \quad k = l+m+1, \ldots, N \end{aligned} \tag{15.20}$$

The coefficients ε^2 and ε determine the rates of change of the variables x. Indeed, the system (15.20) may be rewritten in the form

$$\dot{x}_i = \tau_1^{-1} \mathcal{F}_i, \quad \dot{x}_j = \tau_2^{-1} \mathcal{F}_j, \quad \dot{x}_k = \tau_3^{-1} \mathcal{F}_k$$

where $\tau_1 = \varepsilon^2$, $\tau_2 = \varepsilon$, $\tau_3 = 1$, i.e., $\tau_1 \ll \tau_2 \ll \tau_3$. If we are interested in the behaviour of the system both within small time intervals of the order of τ_1 and within large time intervals of the order of τ_3, we have then to study the system as a whole. The situation is different in other cases. For example, if we are interested in phenomena that occur during average time intervals of the order of τ_2, the system is simplified. Indeed, during such time intervals the "slow" variables x_k have no chance to undergo a change and may be replaced by their initial constant values. On the contrary, the "fast" variables x_i have time to reach their steady-state values, if there are such values. Thus, the variables x_i may be replaced by their steady-state values, and out of N equations there remain only m equations. We shall repeatedly make use of this simplifying procedure in the subsequent discussion.

15.3. Lotka-Volterra Models

The physico-mathematical modelling of biological processes started with the model of an autocatalytic chemical reaction proposed by Lotka (1920) and the prey-predator model advanced by Volterra (1930). These models share many similarities. They belong to the simplest models that describe the behaviour of nonlinear systems and are at the same time important for a number of chemical and biological problems. We shall begin our consideration with these models, which provide us with the required examples of nonlinear dynamics.

Lotka considered the following system of reactions:

$$\mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{X} \tag{I}$$

$$\mathrm{X} + \mathrm{Y} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} 2\mathrm{Y} \tag{II}$$

$$\mathrm{Y} \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{E} \tag{III}$$

The overall reaction is $\mathrm{A} \rightleftarrows \mathrm{E}$. Step (II) is autocatalytic. Substance A is in excess and therefore reaction (I) is zero-order.

The total affinity corresponding to the overall reaction (cf. page 322) is given by

$$\mathcal{A} = \mathcal{A}_{\mathrm{I}} + \mathcal{A}_{\mathrm{II}} + \mathcal{A}_{\mathrm{III}} = RT \ln \frac{k_1 k_2 k_3 [\mathrm{A}]}{k_{-1} k_{-2} k_{-3} [\mathrm{E}]} \tag{15.21}$$

Near equilibrium the system is linear:

$$v_{\mathrm{I}}^0 = k_1 \frac{\mathcal{A}_{\mathrm{I}}}{RT}, \quad v_{\mathrm{II}}^0 = k_2 X_{\mathrm{eq}} Y_{\mathrm{eq}} \frac{\mathcal{A}_{\mathrm{II}}}{RT}, \quad v_{\mathrm{III}}^0 = k_3 Y_{\mathrm{eq}} \frac{\mathcal{A}_{\mathrm{III}}}{RT} \tag{15.22}$$

The stability condition assumes the form

$$k_1 (\delta \mathcal{A}_I)^2 + k_2 X_{eq} Y_{eq} (\delta \mathcal{A}_{II})^2 + k_3 Y_{eq} (\delta \mathcal{A}_{III})^2 \geqslant 0 \qquad (15.23)$$

Near equilibrium it is fulfilled (cf. page 343). Far from equilibrium the reverse reactions may be neglected, i.e., we may put $k_{-1} = k_{-2} = k_{-3} = 0$. Hence, $\mathcal{A} \to \infty$. The kinetic equations are nonlinear and asymmetric:

$$\dot{X} = k_1 - k_2 XY, \quad \dot{Y} = k_2 XY - k_3 Y, \quad \dot{E} = k_3 Y \qquad (15.24)$$

The only steady-state solution corresponds to the condition $\dot{X} = \dot{Y} = 0$. We have

$$X^0 = k_3/k_2, \quad Y^0 = k_1/k_3$$

How will the system behave far from equilibrium? Let us consider this on the basis of the prey-predator model studied by Volterra.

Suppose a certain habitat is shared by predators and their preys, say, lynxes and hares. The lynxes feed only on hares and the latter feed on vegetation abundantly available in the area. The number of preys is X and the number of predators is Y. The behaviour of the populations in time is described by the equations

$$\dot{X} = k_1 X - kXY, \quad \dot{Y} = k'XY - k_2 Y \qquad (15.25)$$

The constant k_1 characterizes the growth of the population of preys; the constant k is their death rate as a result of encounters with predators; k' characterizes the growth of the population of predators, which requires a food supply, i.e., encounters with preys; k_2 is the rate of death of predators. Equations (15.25) are similar in general outline to Eqs. (15.24). Let us find the steady-state values of the variables from Eqs. (15.25) at $\dot{X} = \dot{Y} = 0$:

$$X^0 = k_2/k', \quad Y^0 = k_2/k$$

The quantities X and Y are represented in the form

$$X = X^0 e^{\alpha}, \quad Y = Y^0 e^{\beta} \qquad (15.26)$$

where $\alpha = \ln (X/X^0)$ and $\beta = \ln (Y/Y^0)$ are the measures of the deviation of X and Y from the steady state, in which $\alpha = \beta = 0$. Equations (15.25) will now be rewritten as follows:

$$\frac{1}{k} \dot{\alpha} = Y^0 (1 - e^{\beta}), \quad \frac{1}{k'} \dot{\beta} = -X^0 (1 - e^{\alpha}) \qquad (15.27)$$

We multiply the first equation by $X^0(1-e^{\alpha})$ and the second by $Y^0(1-e^{\beta})$ and then add them up:

$$\frac{1}{k}X^0\dot{\alpha}(1-e^{\alpha})+\frac{1}{k'}Y^0\dot{\beta}(1-e^{\beta})=0$$

or

$$\frac{d}{dt}\left[\frac{1}{k}X^0(e^{\alpha}-\alpha)+\frac{1}{k'}Y^0(e^{\beta}-\beta)\right]=0$$

that is,

$$\frac{1}{k}X^0(e^{\alpha}-\alpha)+\frac{1}{k'}Y^0(e^{\beta}-\beta)=K=\text{const} \qquad (15.28)$$

The quantity K is the constant of motion. Both terms in K are positive since if $\alpha>0$, then $e^{\alpha}>\alpha$, and if $\alpha<0$, then $e^{\alpha}>0$.

We introduce the notations $e^{\alpha}=\varphi$ and $e^{\beta}=\psi$. Multiplying both sides of Eq. (15.28) by kk', we have (since $k'X^0=k_2$ and $kY^0=k_1$):

$$k_2(\varphi-\ln\varphi)+k_1(\psi-\ln\psi)=Kkk'=\text{const}$$

Dividing by k_1k_2, we find

$$\frac{1}{k_1}(\varphi-\ln\varphi)+\frac{1}{k_2}(\psi-\ln\psi)=K\frac{kk'}{k_1k_2}=\text{const}$$

Linearizing, we obtain

$$(\varphi e^{-\varphi})^{1/k_1}(\psi e^{-\psi})^{1/k_2}\equiv UV=\text{const} \qquad (15.29)$$

The equation $UV=\text{const}$ is the equation of the hyperbola. Its plot is shown in Fig. 15.8*a*. Figure 15.8 (*b* and *c*) also shows the behaviour of the functions $V(\psi)$ and $U(\varphi)$. These curves have maxima. The dependence of φ on ψ which follows from Eq. (15.29) is given in Fig. 15.8*d*. To the maxima of U and V there correspond the points A and B on the hyperbola. Upon motion between these two limiting points in the φ, ψ-plane there is described a limit cycle, whose shape depends on the initial conditions. To the steady state there correspond the points A and B on the hyperbola, the points of the maxima M_1 and M_2 on the $U(\varphi)$ and $V(\psi)$ curves and a singular point of the type of the centre in the φ, ψ-plane.

Let us define the behaviour of the system near the singular point. We linearize system (15.25), i.e., we look for its solution of the form

$$X=X^0+xe^{\lambda t},\quad Y=Y^0+ye^{\lambda t} \qquad (15.30)$$

with $|x|, |y| \ll X^0, Y^0$. Neglecting the terms that are nonlinear with respect to x, y, we obtain

$$\lambda x+\frac{kk_2}{k'}y=0,\quad -\frac{k'k_1}{k}x+\lambda y=0 \qquad (15.31)$$

The condition for the compatibility of these two equations is the characteristic equation

$$\lambda^2 + k_1 k_2 = 0 \tag{15.32}$$

and $\lambda = i\omega$ is an imaginary quantity. The oscillation frequency is given by

$$\nu = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{k_1 k_2} \tag{15.33}$$

It is independent of the constants k and k'. Conversely, the amplitudes of oscillations depend on all the four constants. From Eqs. (15.31) and (15.32) it follows that

$$\frac{|x|}{|y|} = \frac{k}{k'}\sqrt{\frac{k_2}{k_1}} \tag{15.34}$$

Thus, the populations X and Y undergo periodic oscillations of equal frequency shifted in phase. For comparison, Fig. 15.9 shows

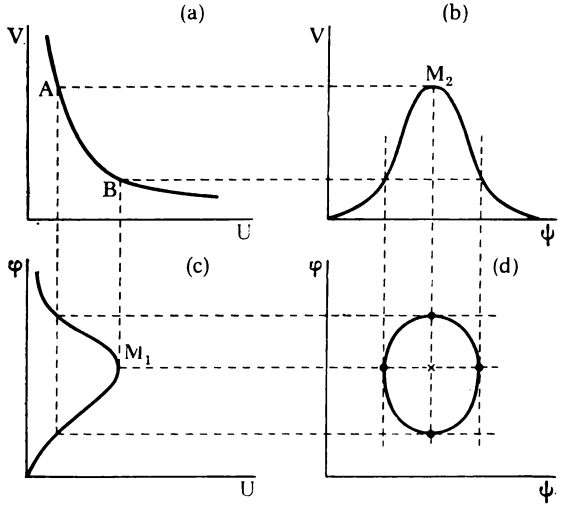


Fig. 15.8. Construction of integral curves for the Lotka-Volterra model. See the explanation in the text.

the dynamics of the lynx and hare populations in Canada for the period 1845-1935 according to the data of a fur company. Such a behaviour also arises from spectacular considerations. Suppose that at the beginning of the period the population of hares was large.

The lynx population increased but the number of hares was reduced. This was followed by a decrease in the number of lynxes, and the size of the hare population increased again, etc. In simple cases, the results of the application of a mathematical model are qualitatively obvious. However, without analysis, it is impossible to establish either the frequency or the amplitude of oscillations of X and Y.

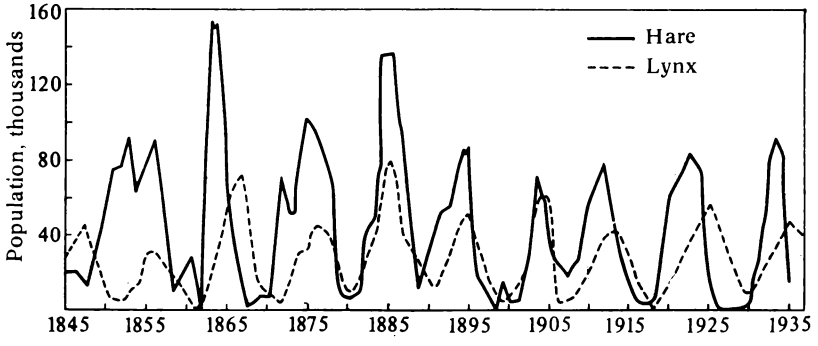


Fig. 15.9. The dynamics of hare and lynx populations in Canada for the period 1845-1935.

In more complicated cases, mathematical models disclose peculiarities that are not obvious at all in advance.

We have examined the behaviour of a conservative system. In fact, the populations of species may increase only up to a certain point, to the values $X^{(m)}$ and $Y^{(m)}$ corresponding to saturation. The maximum of prey reproduction is expressed by the following relation (Verhulst):

$$\dot{X} = k_1 X \frac{X^{(m)} - X}{X^{(m)}} - kXY \qquad (15.35)$$

We have, as before,

$$\dot{Y} = k'XY - k_2 Y \qquad (15.36)$$

The system (15.35), (15.36) is no longer conservative and has no constant of motion.

Equations of the type (15.25) are extended to any number of interacting populations. The study of the corresponding systems is important for ecology.

The Lotka autocatalysis equations, which are similar to the Volterra equations, also lead to a periodic behaviour. We see that nonlinear asymmetric systems are capable of a time-ordered behaviour.

We have already seen that for chemical processes

$$d_X\sigma = \frac{1}{T}\sum v_i d\mathcal{A}_i \leqslant 0 \tag{15.37}$$

(compare with Eq. (15.3)). Near the steady state

$$T\delta_X\sigma = \sum_i \delta v_i d(\delta\mathcal{A}_i) \leqslant 0 \tag{15.38}$$

Expanding δv_i in a series in $\delta\mathcal{A}_i$, we obtain

$$\delta v_i = \sum_j L'_{ij}\delta\mathcal{A}_j \tag{15.39}$$

where $L'_{ij} = (\partial v_i/\partial\mathcal{A}_j)_0$ are derivatives taken for the steady state. Generally speaking, L'_{ij} may contain both a symmetric and an asymmetric part. Under linearity conditions, i.e., near equilibrium, the coefficients L'_{ij} coincide with the Onsager coefficients and their antisymmetric parts are equal to zero. Far from equilibrium this is also possible. Then

$$Td_X\sigma = d\Psi \leqslant 0 \tag{15.40}$$

where

$$\Psi = \frac{1}{2}\sum_{i,j} L'_{ij}\delta\mathcal{A}_i\delta\mathcal{A}_j \tag{15.41}$$

is the kinetic potential. However, if the antisymmetric parts of L'_{ij} are different from zero, then in the general case no kinetic potential exists. Suppose that two chemical processes are described by the antisymmetric matrix L'_{ij}, i.e.,

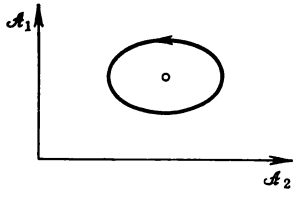


Fig. 15.10. The rotation of an antisymmetric chemical system around the steady state.

$$L'_{11} = L'_{22} = 0, \quad L'_{12} = -L'_{21} = -L'$$

Then

$$T\delta_X\sigma = L'(\delta\mathcal{A}_2 d(\delta\mathcal{A}_1) - \delta\mathcal{A}_1 d(\delta\mathcal{A}_2)) \leqslant 0 \tag{15.42}$$

Such a system rotates around the steady state, without getting into it. The phase pattern has the form shown in Fig. 15.10; the point corresponding to the steady state is the centre. Introducing the polar coordinates r and φ in the $\mathcal{A}_1$, $\mathcal{A}_2$ plane, we obtain

$$Td_X\sigma = -L'r^2 d\varphi \leqslant 0 \tag{15.43}$$

But the function $\Psi = L'r^2\varphi$ is not a potential—it increases by $2\pi L'r^2$ with every revolution.

It is exactly this situation that has been dealt with in considering the Lotka-Volterra system.

15.4. Autocatalytic Systems

The Lotka chemical autocatalytic system considered above is an oscillatory system: it is characterized by a singular point of the centre type. Let us now discuss more complex point autocatalytic chemical systems. We shall consider the so-called *Brusselator*—the Prigogine-Lefever-Nicolis model.

We have four coupled reactions:

$$A \xrightarrow{k_1} X \tag{I}$$

$$2X + Y \xrightarrow{k_2} 3X \tag{II}$$

$$B + X \xrightarrow{k_3} D + Y \tag{III}$$

$$X \xrightarrow{k_4} E \tag{IV}$$

The overall reaction is A + B → D + E. Reaction (II) is autocatalytic.

The kinetic equations for the intermediate compounds X and Y are as follows:

$$\begin{aligned} \dot{X} &= k_1 A + k_2 X^2 Y - k_3 B X - k_4 X \\ \dot{Y} &= -k_2 X^2 Y + k_3 B X \end{aligned} \tag{15.44}$$

We introduce the dimensionless variables

$$\tau = k_4 t, \quad x = X\sqrt{k_2/k_4}, \quad y = Y\sqrt{k_2/k_4}$$
$$a = A\,(k_1/k_4)\sqrt{k_2/k_4}, \quad b = Bk_3/k_4$$

The following equations are obtained:

$$\begin{aligned} dx/d\tau &= a + x^2 y - bx - x \\ dy/d\tau &= -x^2 y + bx \end{aligned} \tag{15.45}$$

Analogous equations are obtained at $k_1 = k_2 = k_3 = k_4 = 1$ and $\tau = t$. We find the steady-state solution by equating the left-hand sides of Eqs. (15.45) to zero. From the second equation we find $x_0^2 y_0 = bx_0$; substituting into the first equation, we obtain $x_0 = a$ and, hence, $y_0 = b/a$. This solution is the only one. We study its stability. We seek the solution for Eqs. (15.45) in the form

$$x = a + \alpha e^{\lambda\tau}, \quad y = b/a + \beta e^{\lambda\tau} \tag{15.46}$$

with $|\alpha|, |\beta| \ll a, b/a$. In the linear approximation we get

$$\alpha\lambda = (b-1)\alpha + a^2\beta, \quad \beta\lambda = -b\alpha - a^2\beta$$

and the characteristic equation

$$\lambda^2 + (a^2 + 1 - b)\lambda + a^2 = 0 \tag{15.47}$$

which has the roots

$$\lambda_{1,2} = -\frac{1}{2}(a^2 + 1 - b) \pm \frac{1}{2}\sqrt{(a^2 + 1 - b)^2 - 4a^2} \tag{15.48}$$

Since $\lambda_1\lambda_2 = a^2 > 0$, it follows that the roots are either real and have equal signs or are complex-conjugate. Hence, we deal with cases 1, 2 or 5, 6 of the classification given on page 510. The singular

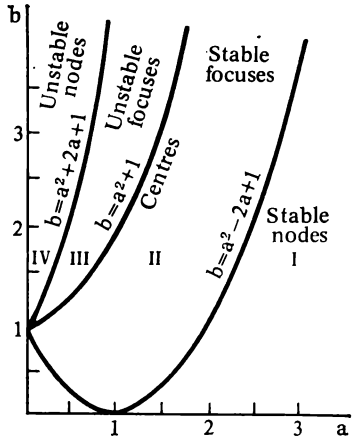


Fig. 15.11. The regions in the a, b plane.

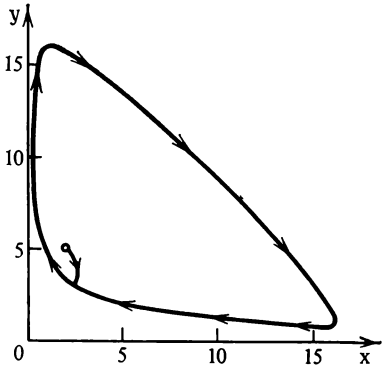


Fig. 15.12. The phase trajectory of the Brusselator.

point corresponding to the steady state is a stable or unstable node or focus. The system becomes unstable when the parameter b exceeds the value that satisfies the condition

$$p = a^2 + 1 - b = 0$$

The value $b_{cr} = a^2 + 1$ is the critical, bifurcation point. The parabola $b = a^2 + 1$ in the a, b-plane separates the stable regions from the unstable ones: the system has unstable nodes or foci at $b > b_{cr} = a^2 + 1$ and stable nodes and foci at $b < b_{cr} = a^2 + 1$. To the parabola itself, $b = a^2 + 1$, there correspond the centres—at this value $\lambda_{1,2} = \pm ai$ and the system performs undamped oscillations. The curves that separate the focal region from the nodal

region are found from the condition of the equality of the square root in Eq. (15.48) to zero. We obtain

$$b = (a \pm 1)^2$$

Thus, the a, b-plane is divided up into four regions (Fig. 15.11). In region I the deviations from the steady state decrease exponentially; in region IV they increase exponentially. In region II damped oscillations occur, and in region III the occurrence of unstable auto-oscillations is observed. The exponential increase in region IV at a large amplitude is interrupted and repeats periodically—a limit cycle occurs. The phase trajectory of the Brusselator at $a = 2$ and $b = 10$, i.e., in region IV, is shown in Fig. 15.12. The limit cycle has a shape close to a triangle and the auto-oscillations are of a sharp relaxational character (Fig. 15.13). It should be emphasized that the limit cycle is not considered in the study of linearized equations—it emerges at a sufficient distance from the singular point. The results presented in Figs. 15.12 and 15.13 have been obtained by means of a numerical integration of nonlinear equations.

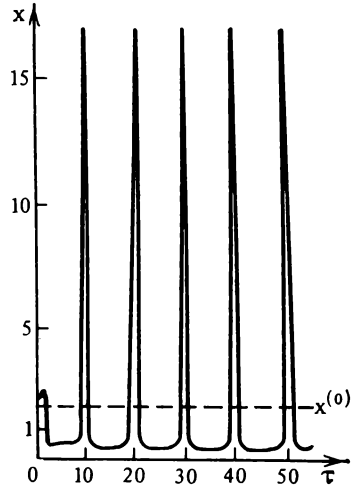


Fig. 15.13. The auto-oscillations of the Brusselator.

The Brusselator differs from the Lotka-Volterra system. The latter has an infinite number of possible periodic orbits around the steady state, the transitions between which are accomplished through changes in the initial conditions and the parameters. Conversely, the Brusselator is a nonconservative system which comes to a time-ordered coherent auto-oscillatory regime, independently of the initial conditions if the system's parameters correspond to the region of unstable nodes.

Thus, point autocatalytic systems, one example of which is the Brusselator, are capable of passing over to a state ordered with respect to time.

We now turn to *distributed autocatalytic systems.* As we shall see, far from equilibrium such systems can form space-ordered dissipative structures. The distribution of the system in space is determined by the presence of diffusion.

The problem of the stability towards diffusion was first studied by Turing (1952) in a remarkable paper "On the Chemical Basis of Morphogenesis". The ultimate goal of this work and also of the investigations based on it is to arrive at a model interpretation of biological phenomena that determine the ontogenetic development (see Chapter 17).

Let us supplement Eqs. (15.45) by terms that describe one-dimensional diffusion along the coordinate r. We have

$$\begin{aligned} \frac{\partial x}{\partial \tau} &= a + x^2 y - bx - x + D_x \frac{\partial^2 x}{\partial r^2} \\ \frac{\partial y}{\partial \tau} &= -x^2 y + bx + D_y \frac{\partial^2 y}{\partial r^2} \end{aligned} \tag{15.49}$$

The substances x and y (i.e., X and Y) are characterized by different one-dimensional-diffusion constants, D_x and D_y.

We look for the solutions of the system (15.49) in the form of concentration waves (cf. Eqs. (15.46)):

$$\begin{aligned} x &= x_0 + \alpha \exp(\lambda\tau + ir/l) \\ y &= y_0 + \beta \exp(\lambda\tau + ir/l) \end{aligned} \tag{15.50}$$

Here l is the wavelength which characterizes the space inhomogeneity, $\alpha \ll x_0$, $\beta \ll y_0$, $x_0 = a$, $y_0 = b/a$. Substituting Eqs. (15.50) into Eqs. (15.49), we arrive at the characteristic equation

$$\lambda^2 + (a^2 + 1 - b + \xi + \eta)\lambda + a^2(1 + \xi) + (1 - b)\eta + \xi\eta = 0 \tag{15.51}$$

where $\xi = D_x/l$, $\eta = D_y/l$. At $l \to \infty$ the system becomes homogeneous and Eq. (15.51) coincides with Eq. (15.48).

Equation (15.51) gives two types of instability. The transition to the first type occurs when the factor at λ is equal to zero and the sum of the free terms is positive:

$$b'_{cr} = 1 + a^2 + \xi + \eta$$

The second transition corresponds to the case where the sum of the free terms is equal to zero:

$$b''_{cr} = (a^2 + \eta)(1 + \xi)/\eta$$

In the second case one of the roots in Eq. (15.51) vanishes. Thus, the solutions are stable in the region

$$b'_{cr} < b < b''_{cr}$$

The instability conditions have the form

$$b > b''_{cr} \quad \text{and} \quad b < b'_{cr}$$

The wavelength l at which b''_{cr} is minimal is $l \to \infty$, $\xi, \eta \to 0$ and min $b'_{cr} = 1 + a^2$. The minimum of b''_{cr} is found by equating the derivative of b''_{cr} with respect to l to zero. We obtain $l^2_{cr} = (D_x D_y)^{1/2}/a^{-1}$ and min $b''_{cr} = [1 + a(D_x/D_y)^{1/2}]^2$. Instability arises when b reaches the lesser of these two values. If $D_x = D_y$, then

$$\min b'_{cr} = 1 + a^2 < \min b''_{cr} = (1 + a)^2$$

and a limit cycle is realized. But at sufficiently small D_x/D_y, min $b'_{cr} >$ min b''_{cr} —there appears space inhomogeneity, i.e., a symmetry-breaking occurs. For b larger than b_{cr} the fluctuations with l close to l_{cr} are amplified and the system leaves the steady state. Then it is stabilized in a new steady state, which is spatially

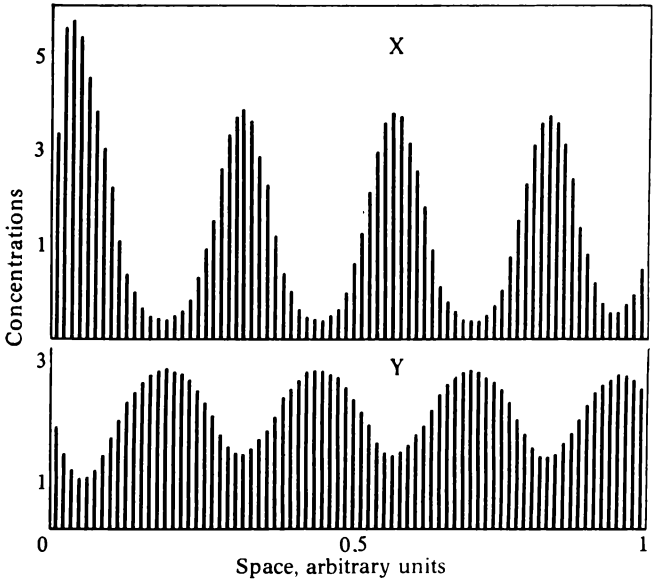


Fig. 15.14. The steady-state distribution of a substance as a result of a symmetry breaking instability. The concentrations X and Y on the boundaries of the region are kept equal to the homogeneous steady-state values of $X = 2$ and $Y = 2.62$. The values of the parameters are as follows:

$a = 2.000$, $b = 5.24$, $D_x = 1.6 \cdot 10^{-3}$, $D_y = 8 \cdot 10^{-3}$, i.e., min $b'_{cr} \approx 5$ and min $b''_{cr} \approx 7.7$.

inhomogeneous. Figure 15.14 shows the calculated steady-state nonuniform distribution of X and Y in space for the system (15.49) which arises beyond the limits of a symmetry-breaking instability.

In considering the point and distributed Brusselators we neglected the reverse reactions and, thereby, moved far away from equilibrium. The spatial structure depicted in Fig. 15.15 is stabilized by a flux of energy and matter passing through an open system. This dissipative structure is characterized by a lower entropy than the original homogeneous system—a phenomenon called *order through fluctuations* is realized.

Of course, a real system is not a homogeneous system with constant boundary conditions. The concentrations themselves (dimensionless), a and b, that appear as the parameters, depend on r and t.

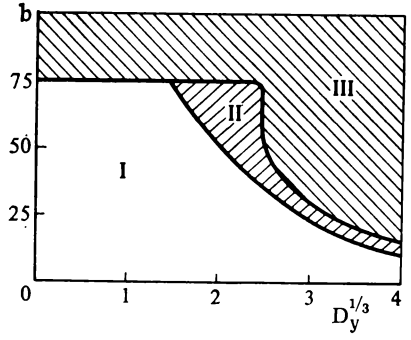


Fig. 15.15. The properties of the autocatalytic system under consideration at given values of a, D_a, D_x and $D_b \to \infty$ as functions of b and D_y (in arbitrary units):
I—region of stable steady state; *II*—region of a monotonic increase of fluctuations; *III*—region of amplified oscillations.

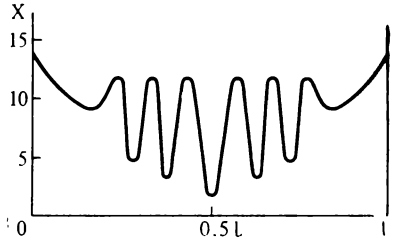


Fig. 15.16. A localized steady-state dissipative structure that arises in region *II* of Fig. 15.15.

Let us consider the same system (15.49), assuming for simplicity that b is constant and taking account of the changes of a:

$$\frac{\partial a(r, t)}{\partial t} = -a(r, t) + D_a \frac{\partial^2 a(r, t)}{\partial r^2} \qquad (15.52)$$

We look for a general solution of Eqs. (15.49) and (15.52) that would satisfy the boundary conditions

$$a(0) = a(L) = \bar{a}, \quad x(0) = x(L) = \bar{x}$$
$$y(0) = y(L) = \bar{y} \quad (0 \leqslant r \leqslant L)$$

The properties of this solution are as follows:

1. Near equilibrium there is only one stable steady state.
2. Far from equilibrium various situations may arise (Fig. 15.15).

3. In region *II* of the diagram (Fig. 15.15) the thermodynamic (static) branch becomes unstable with respect to fluctuations in chemical composition. Beyond the threshold of b the initial deviations from the steady state become larger and eventually bring the system to a new steady state corresponding to the nonuniform distribution of x and y (Fig. 15.16).

4. For $D_b \gg D_a \gg D_y \approx D_x$ the thermodynamic branch is unstable and the system passes over to a spatially inhomogeneous time-dependent state. There arise *concentration waves* which propagate in the reaction volume and are reflected from the boundaries of the system. At each point there occur oscillations in concentration.

5. At finite values of D_a the structures described are localized inside the reaction volume. Their boundaries and also periods are determined by the values of the quantities a, b, D_x, etc., independently of the initial conditions. At D_a, $D_b \to \infty$ the structures occupy the entire volume. If D_x and D_y are very large as compared with the chemical rates in region *III* (Fig. 15.15), the space dependence disappears and the system oscillates everywhere with the same phase. There is realized a limit cycle and the periodic motion is stable.

The concentration oscillations and waves will be discussed in Sec. 16.1.

15.5. Phase Transitions

The emergence of static order from disorder under thermodynamic equilibrium conditions occurs through a phase transition. In the transition region the system is unstable. In a first-order transition, say, in the transition gas → liquid, the basic thermodynamic quantities—enthalpy, entropy, and volume—undergo a breakdown. In a second-order phase transition the state of the system is continuously altered, but symmetry is changed in a discontinuous manner. In this case, not the basic thermodynamic quantities but their derivatives are broken—heat capacity, compressibility, and the thermal expansion coefficient. An example is the transition paramagnetic → ferromagnetic at the Curie point.

The points of phase transitions are singular points at which the character of the functional dependence of the chemical potential on temperature and pressure is changed.

As we have already seen, transitions involving the appearance of space and/or time order far from equilibrium also occur discontinuously, being associated with the instabilities of the initial states. The transition of radiating atoms at the critical values of the optical pumping parameter to a coherent laser radiation exhibits all the specific features of a phase transition.

Consider a simpler physical example—an electric circuit with a diode (Fig. 15.17). The straight dotted line on the graph represents

a current that flows through a resistor as a function of the voltage on the diode, and the solid line represents the current-voltage characteristic of the diode. Under steady-state conditions the currents that pass through the resistors and the diode must be equal—this is marked by the points A, B, and C. In order to determine the stabilities of these states, it is necessary, as is always the case, to linearize equations that describe the circuit near those states. The linearized circuit contains a capacitor connected in parallel with

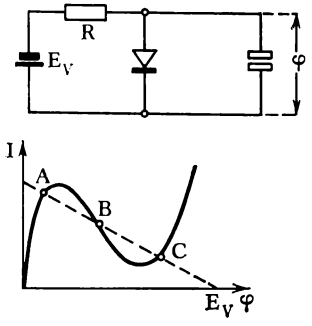


Fig. 15.17. An electric circuit with a diode and its current-voltage characteristic

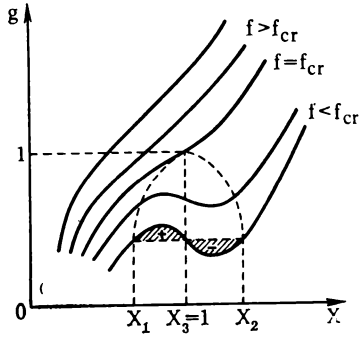


Fig. 15.18. A first-order phase transition in an autocatalytic chemical reaction.

the resistor R and the differential resistor $d\varphi/dI$ of the diode. If these two parallel resistances are positive, the circuit is stable. This is valid for the points A and C but not for the point B. The system has two different stable steady states, A and C, it is bistable, and the transition from one state to the other has the character of a phase transition. The system is a flow, nonequilibrium system.

We shall now consider autocatalytic chemical reactions (Schlögl, 1972). We have already seen that chemical processes of this kind can give rise to dynamic order.

Consider the reactions

$$A + 2X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 3X \qquad \text{(I)}$$

$$B + X \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C \qquad \text{(II)}$$

The overall reaction is $A + B \rightleftharpoons C$. The reaction rates are given by

$$v_1 = k_1 AX^2 - k_{-1}X^3 \quad \text{and} \quad v_2 = k_2 BX - k_{-2}C \qquad (15.53)$$

We put for simplicity $k_{-1} = 1$ and $k_1 A = 3$ and introduce the notations $k_2 B = f$ and $k_{-2} C = g$. The rate of change of the concentration X is

$$\dot{X} = v_1 - v_2 = -X^3 + 3X^2 - fX + g \tag{15.54}$$

In the steady states $\dot{X} = 0$ or

$$g = X^3 - 3X^2 + fX \tag{15.55}$$

The $g(X)$ curves, which represent this dependence, are shown in Fig. 15.18. The equation $\dot{X} = 0$ has three roots. These roots coincide at the critical value of $f = f_{cr}$:

$$g = 1, \quad f_{cr} = 3, \quad X_{cr} = 1$$

Three different real and positive values of the roots $X_2 > X_3 > X_1$ are possible only at $f < 3$. The solutions of X_1, X_2 are stable and the unstable root X_3 is on the descending branch of the $g(X)$ curve. The picture obtained is similar to that for the circuit with a diode. Even more informative is the similarity to the gas-liquid phase diagram. Here the concentration X plays the role of the density V^{-1}; the quantity g corresponds to the pressure p and the quantity f to the temperature RT. We obtain

$$g = \frac{f}{V} - \frac{3}{V^2} + \frac{1}{V^3} \tag{15.55a}$$

The van der Waals equation has the form

$$p = \frac{RT}{V-b} - \frac{a}{V^2} \tag{15.56}$$

or, expanding into a series in b (the virial form),

$$p = \frac{RT}{V} + \frac{bRT - a}{V^2} + \frac{b^2 RT}{V^3} + \ldots \tag{15.57}$$

A full analogy obtains for the virial equation of state

$$p = \frac{RT}{V} - \frac{a_1}{V^2} + \frac{a_2}{V^3} \tag{15.58}$$

with the coefficients a_1 and a_2 being independent of temperature.

Thus, in the autocatalytic system under consideration there occurs, at $f < f_{cr}$, a transition between two stable steady states X_1 and X_2 which is similar to a first-order phase transition. To the external conditions specified by f and g there correspond two competing states, X_1 and X_2. This transition is not associated with symmetry breaking (a gas and a liquid have the same symmetry). It can be shown that the Maxwell condition—the equality of the areas + and — in Fig. 15.18 is fulfilled.

If the chemical reaction is accompanied by the diffusion of substance X (the system is a distributed system), then at constant values of A, B, and C we obtain

$$\dot{X} = g(X) + D\frac{\partial^2 X}{\partial r^2} \tag{15.59}$$

Here the dependence $g(X)$ may be represented in the form

$$g(X) = \frac{\partial}{\partial X}\Phi(X) \equiv \frac{\partial}{\partial X}\left(\frac{1}{4}X^4 - X^3 + \frac{1}{2}fX^2\right) \tag{15.60}$$

where $\Phi(X)$ plays the role of a certain potential. The steady states satisfy the condition $\dot{X} = 0$, i.e., the equation

$$D\frac{\partial^2 X}{\partial r^2} = -\frac{\partial \Phi}{\partial X} \tag{15.61}$$

This equation is similar to the equation that describes the motion of a material point in a potential field if D is the "mass", r is the "time", and X is the "coordinate". There exists a steady state in

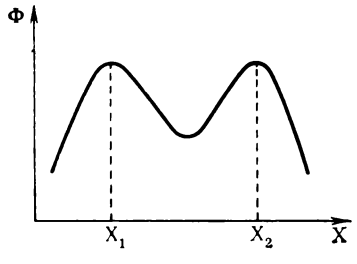


Fig. 15.19. The coexistence of phases X_1 and X_2.

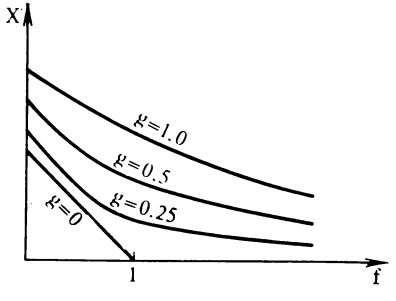


Fig. 15.20. A second-order phase transition in an autocatalytic chemical reaction.

which two phases, X_1 and X_2, coexist, with $X \to X_1$ at $r \to \infty$ and $X \to X_2$ at $r \to -\infty$; to the values X_1 and X_2 there correspond the relative maxima of $\Phi(X)$. In the steady state the "mass" D is shifted from one maximum of $\Phi(X)$ at which this "mass" was at rest to the other maximum at which it also reaches a state of rest. This is possible only if the values of Φ at both points are equal (Fig. 15.19):

$$\Phi(X_1) = \Phi(X_2) \tag{15.62}$$

This is the condition for the coexistence of two "phases" that fixes the value of g. It can be shown that this condition leads to

$$g = f - 2,\ X_{1,2} = 1 \mp (3-f)^{1/2},\ X_3 = 1 \tag{15.63}$$

Apart from first-order transitions, nonlinear autocatalytic reactions can lead to nonequilibrium transitions similar to second-order phase transitions. Consider the reactions

$$A + X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2X \tag{I}$$

$$B + X \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C \tag{II}$$

Their rates are equal to

$$v_1 = k_1 AX - k_{-1} X^2,\quad v_2 = k_2 BX - k_{-2} C \tag{15.64}$$

Putting $k_1 A = 1$, $k_{-1} = 1$, $k_2 B = f$ and $k_{-2} C = g$, we find that

$$\dot{X} = v_1 - v_2 = -X^2 + (1-f)\,X + g \equiv \varphi(X) + g \tag{15.65}$$

Here $\dot{X} > 0$ at small X and $\dot{X} < 0$ at large X. Hence, the steady state corresponding to $\dot{X} = 0$ is stable. In this state

$$g = X^2 - (1-f)\,X \equiv -\varphi(X) \tag{15.66}$$

If $g = 0$, then

$$X = \begin{cases} 1-f & \text{at } f < 1 \\ 0 & \text{at } f \geqslant 1 \end{cases} \tag{15.67}$$

Such a behaviour of the quantity X is characteristic of second-order transitions, say, the transition from the ferromagnetic to the paramagnetic state. In this analogy X plays the role of the order (magnetization) parameter; g is the magnetic field strength; f is the temperature. The value of $f = f_{cr} = 1$ corresponds to the Curie point. At $f < f_{cr}$ the symmetry is broken. It is exactly at the critical point that small fluctuations increase up to macroscopic values. Figure 15.20 shows curves of X vs. f for different values of g.

The behaviour of a model chemical system which is similar to a phase transition implies its collective properties. The transitions occur according to the all-or-none law. They take place in trigger systems as a result of small fluctuations.

The diverse biological phenomena are of the dynamical nature similar to that of phase transitions. The further development of the theory of nonequilibrium phase transitions as the mechanisms of biological processes seems to be promising. Some interesting results have already been obtained. Chaikovsky and Ebeling (1977) have studied such transitions in systems of enzymatic reactions in the

presence of substrate inhibition and diffusion. They have demonstrated the possibility of the spatial separation of phases corresponding to complexes of an enzyme with several substrate molecules. Belintsev and Volkenshtein (1977) have shown that the emergence of a new species as a result of the geographic separation of populations is similar to a phase transition. Similar phenomena occur in a number of membrane processes, in particular, in the generation of a nerve impulse (see Sections 11.3 and 16.5).

15.6. Stochastic Processes

In describing dynamical processes we employed the apparatus of differential equations. But, as has been said (page 504), with a limited number of molecules or species one has to take direct account of the discreteness of their emergence and destruction—the acts of "birth" and "death". No half molecule or half hare can appear or disappear in nature. Discrete systems are described by stochastic equations based on probability theory. We shall confine ourselves to a consideration of two examples.

Let us return to the system of chemical reactions considered on page 531:

$$A+X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2X \qquad (I)$$

$$B+X \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C \qquad (II)$$

Considering that the molecules are "born" and "die", we derive the *master equation* for the system, the variables being here not the concentrations but the numbers of molecules. Let N denote the number of molecules of X; the probability distribution, which changes with time, is expressed by the function $P(N, t)$. In each of the reaction N is changed by unity. Let us begin with reaction (II).

1. The transition $N \to N + 1$ is the "birth" of a molecule of X. The number of such transitions per unit time is equal to $P(N, t)\, W(N + 1, N)$, where $W(N + 1, N)$ is the transition probability given by

$$W(N + 1, N) = k_{-2}CV$$

where C is the concentration of molecules of C, i.e., their number in unit volume; V is the volume; CV is the number of reacting molecules of C.

2. The transition $N - 1 \to N$ is also the "birth" of X. The number of such transitions per second is $P(N - 1, t)\, k_{-2}CV$. Taking into account the fact that the "population" of the level N is simul-

taneously diminished as a result of process (I), we obtain the total rate of transition in the reaction $C \rightarrow B + X$:

$$[P(N-1, t) - P(N, t)]\, k_{-2}CV$$

Let us now consider the reaction $B + X \rightarrow C$. In this reaction the molecules of X "die"; the number N decreases by unity. We have the transitions $N+1 \rightarrow N$ and $N \rightarrow N-1$. The total rate is given by

$$[P(N+1, t)(N+1) - P(N, t)N]\, k_2B$$

where B is the concentration of B. In an analogous way we find the rates for the forward and reverse reactions (I). In the reaction $A + X \rightarrow 2X$ the molecules of X are "born" at the rate

$$[P(N-1, t)(N-1) - P(N, t)N]\, k_1A$$

and in the reaction $2X \rightarrow A + X$ they "die" at the rate

$$[P(N+1, t)(N+1)N - P(N, t)N(N-1)]\, k_{-1}/V$$

We can now write the master equation. It has the form

$$\dot{P}(N, t) = P(N-1, t)\, W(N, N-1) + P(N+1, t)\, W(N, N+1) - P(N, t)\,[W(N+1, N) + W(N-1, N)] \quad (15.68)$$

where

$$\begin{aligned} W(N, N-1) &= \left(k_1A\frac{N-1}{V} + k_{-2}C\right)V \\ W(N, N+1) &= \left[k_{-1}\frac{(N+1)N}{V^2} + k_2B\frac{N+1}{V}\right]V \end{aligned} \quad (15.69)$$

There is a difficulty which consists in that the steady-state solution of Eq. (15.68) is $P(0) = 1$, $P(N) = 0$ at $N \neq 0$. Therefore we have to introduce a third process of the spontaneous birth of molecules of X from molecules of A:

$$A \xrightarrow{k_1} X \quad \text{(III)}$$

in which case formulas (15.69) transform to

$$\begin{aligned} W(N, N-1) &= \left(k_1A\frac{N}{V} + k_{-2}C\right)V \\ W(N, N+1) &= \left[k_{-1}\frac{(N+1)N}{V^2} + k_{-2}B\frac{N+1}{V}\right]V \end{aligned} \quad (15.70)$$

The solution of Eq. (15.68) takes the form

$$P(N) = P(0)\prod_{n=0}^{N-1}\frac{W(n+1, n)}{W(n, n+1)} \quad (15.71)$$

In order to pass to the deterministic continuity equation (15.64), (15.65), let us find the average values of N and the probabilities W. We have

$$\langle N\rangle = \sum_{N=0}^{\infty} NP(N, t) \tag{15.72}$$

$$\langle W(N \pm 1, N)\rangle = \sum_{N=0}^{\infty} W(N \pm 1, N) P(N, t) \tag{15.73}$$

and

$$\frac{d}{dt}\langle N\rangle = \langle W(N+1, N)\rangle - \langle W(N-1, N)\rangle \tag{15.74}$$

Using Eqs. (15.69), we obtain

$$\frac{d}{dt}\langle N\rangle = \left[k_1 A \frac{\langle N+1\rangle}{V} + k_{-2}C - k_{-1}\frac{\langle N(N-1)\rangle}{V^2} - k_2 B\frac{\langle N\rangle}{V}\right] V \tag{15.75}$$

This equation coincides with the continuity equation at large N, when unity may be neglected as compared with N and $\langle N\rangle/V$ may be replaced by X. We obtain

$$\dot{X} = k_1 AX + k_{-2}C - k_{-1}X^2 - k_2 BX = v_1 - v_2$$

which coincides with Eq. (15.64). The replacement of $\langle N\rangle/V$ by the concentration is quite strict if P is a Poisson distribution. The condition for such a distribution is the detailed balance for the distributions in each of reactions (I) and (II) separately. This is always valid at thermodynamic equilibrium but may not be fulfilled far from it. Therefore, the question of the relationship between deterministic and stochastic equations always requires a special study.

As a second example, let us consider the prey-predator model proposed by Volterra (Sec. 15.3). Let us denote the number of hares by M and the number of lynxes by N. The reproduction of hares, i.e., the transitions $M \to M + 1$, is described by the probability

$$W(M + 1, N; M, N) = k_1 M$$

The death of lynxes is described by the probability of the transitions $N \to N - 1$:

$$W(M, N - 1; M, N) = k_2 N$$

In encounters between lynxes and hares there occur the transitions $M \to M - 1$ and $N \to N + 1$ with the probabilities

$$W(M - 1, N + 1; M, N) = kMN$$

We put $k' = k$ (see page 516). The basic kinetic equation has the form

$$\dot{P}(M, N; t) = k_1(M-1)P(M-1, N; t) + \\ + k_2(N+1)P(M, N+1; t) + k(M+1)(N-1) \times \\ \times P(M+1, N-1; t) - (k_1 M + k_2 N + kMN)P(M, N; t) \quad (15.76)$$

Of course, at negative values of M and N there is imposed the condition $P = 0$. The only steady-state solution, $\dot{P} = 0$, corresponds to $P(0, 0) = 1$ and the remaining ones to $\dot{P} = 0$. In other words, both species will be exterminated. The comparison with the Lotka-Volterra differential equations discloses the meaning of this result. We have already seen that the closed phase trajectories surrounding the centre can change with any fluctuations—the system moves from one trajectory to another when the initial conditions are altered. If the hares were accidentally exterminated, the lynxes would not survive and $M = N = 0$ is the only steady state. The hares can hide from the lynxes or run away from them to other areas.

We see that the analysis of a stochastic equation yields in this case new informative results.

15.7. Dynamics and Regulation

The emergence of a collective behaviour of an open system ordered in time and space as a result of a transition similar to a phase transition implies the presence of special regulatory potentialities in such systems. The spatial localization of matter (see page 525) is a probable mechanism of stabilization of a dissipative structure with respect to changes in the chemical environment.

Localized systems are characterized by the production of large amounts of a certain substance in a limited region of space. The system thus displays regulatory properties. The concentration waves (page 526) constitute a mechanism of propagation and transfer of information in the form of chemical, molecular signals, and steady-state localized structures store the information.

Biological macromolecules, supermolecular structures, cellular organelles, cells, organisms, populations are all complex systems, i.e., assemblies of elements interacting with one another. The study of life phenomena is based on the investigations of such interactions. At the same time, a physical consideration of a complex system cannot be carried out without studying its constituent elements taken separately, including the molecular level of organization. The interactions themselves are controlled by the nature of these elements. Accordingly, we deal with an enzyme and a gene, with an axon and a myofibril, with a mitochondrion and a chloroplast. These elements

of complicated systems are complex systems themselves. The analysis of life phenomena at all levels of organization requires approaches consistent with the concepts of the *general theory of systems.*

Specific interactions in biological systems lead to the regulation of their behaviour, the maintenance of constant values of the vital parameters of the system under steady-state conditions, and to the directed self-organization of a developing organism.

In biology there has long appeared the concept of *homeostasis*, which implies the tendency to maintain the internal environment at a constant level. However, as has been pointed out by Waddington, one should speak not of homeostasis but rather of *homeorhesis*—life phenomena are dynamical in character and here the stabilization of flows is essential. Homeorhesis implies the presence of a steady state or a stationary closed trajectory on the phase pattern of an open system, in particular, the presence of a limit cycle. Homeorhesis is maintained by regulatory processes, which restore the phase trajectories upon deviations from them that are caused by changes in the environmental conditions.

Using the language of the theory of control, let us say that an open system is characterized by the presence of an input and an output signal—the action on the system and its response to that action. The law of the behaviour of the system determines the dependence of the output quantity on the input influence. The tasks of the theory of systems in their general outline consist in combining two known factors with the purpose of finding the third. These three factors are the input quantity, the behaviorial law, and the output quantity. The key problems of biophysics boil down to finding the structure and laws of the behaviour of a "black box"—a biological system. An enzyme, a cell, and an organism are all "black boxes". The engineer designs a "white box"—a machine that transforms the input signals to the output signals according to the intended goal. In biology, "boxes" have been created by nature and the objective of physics is to study their internal structure and functionality.

Regulation, which ensures the maintenance of homeorhesis or optimal responses to external influences, is accomplished as a result of the interaction between the output and input signals, i.e., as a result of *feedback*. In the simplest case, regulation maintains the output quantity at a constant level, say, the temperature of the laboratory thermostat. The simplest regulatory system contains a controlling device, which is subject to the influence of the output signal, and the object to be controlled, which produces this signal (see page 485).

The interactions in a biological system are either strong (chemical interactions) or weak (intermolecular interactions) (see Sec. 2.11). As a rule, chemical reactions and also weak cooperative interactions are nonlinear. Nonlinearity is necessary for the control system to

be realized. Cooperativity invariably implies the nonlinearity of the response of the system to the input signal.

As we have seen, nonlinear systems behave in diversified ways. The presence of multiple steady states determines the possibility of switching the system over from one regime to another even in the case of weak interactions. Thus, nonlinear systems possess regulatory facilities.

The mathematical apparatus of the theory of control systems consists of differential equations. Such an equation describes the relationship between input and output signals. The method of transfer functions based on the application of Laplace transforms allows one to obtain a phenomenological description of control systems. In such a case, the method of phase patterns described above is efficient; it permits a direct analysis of stability problems.

The consideration of an organism as a controlled system carried out by means of physico-mathematical methods is the basis of theoretical physiology. In Sec. 16.6 we shall consider, as an example, the problem of fibrillation of the cardiac muscle. Theory enables one to establish the parameters that govern the behaviour of the system and to find methods for influencing these parameters. As a matter of fact, a number of physiological problems are of the same nature (in particular, the problem of immunity, Sec. 17.8). It is necessary to unravel the physical essence of regulatory phenomena, to establish what factors govern the regulation and, thereby, to find the causes of pathologic abnormalities. In this sense, physics is the basis of physiology.

Chapter 16 Periodic Chemical and Biological Processes

16.1. Introduction

In the preceding chapter we have dealt with auto-oscillatory and auto-wave processes characteristic of open far-from-equilibrium systems. The only experimental fact that we have resorted to so far is the periodic change of the populations of hares and lynxes corresponding to the Volterra model (page 519). The number of such biological facts is, however, very large. At all levels of organization, from the macromolecular to the population level, in biological systems there occur undamped oscillations of characteristic physical parameters, such as enzyme activity and metabolite concentrations, the parameters that govern the physiological behaviour, the size of populations, etc.

One could offer arguments in favour of the point of view that a biological system not only can but must be oscillatory. The first argument follows from what has been expounded in the preceding chapter. A complex open system that involves autocatalytic chemical reactions often, while being far from equilibrium, evolves toward limit cycles. Consequently, auto-oscillations in such a system are quite probable. The second argument has been given by Molchanov. Biological systems evolved as a result of a very long evolutionary process. During the evolution stable systems must have been equilibrated and become part of the outside world. On the contrary, unstable systems degraded in the course of the evolution. Hence, only those systems, whose internal motions are of the oscillatory nature, could have survived.

The evolution, both prebiotic and biological, has taken place on the Earth which revolves around the Sun and rotates on its axis. This could not but affect the course of the evolution (see Sec. 17.5). At the same time, under the conditions of the periodic change of temperature, light, and humidity the evolution must have reflected in the physiology of both animal and plant organisms. This is the third argument. The day-and-night, circadian periodicity is really characteristic of life processes. In this connection there has been introduced the concept of *biological clocks*. Biological oscillations having a period close to a period of 24 hours, such as, for example, the change of sleep and awakedness, are referred to as *circadian rhythms*. Circadian rhythms are operative everywhere in living nature; they are endoge-

nous, internally generated. The physiological factor responsible for circadian rhythms undoubtedly lies in auto-oscillatory processes—their periods are practically independent of temperature and other environmental factors. Circadian rhythms are autonomous both at the organismic and the cellular level.

The fourth, no less important, argument is that any complex machine can operate continuously only by way of a periodic transformation of energy into work. Periodicity is inherent in any movable device—one of the greatest achievements of *Homo sapiens* was the wheel (see, however, page 434). It is obvious that a complex living system capable of autonomous existence attains, in an evolutionary way, the level of a periodically operating machine—we mean here the respiration and blood-circulation systems. The movements of an animal—the running of a cheetah, the jumps of a kangaroo, the flight of a bird, the swimming of fish, the creeping movement of a snake, the motion of ciliata—are all periodic, often auto-wave oscillatory, processes involving the conversion of chemical energy into mechanical work (see Chapter 12). The theoretical and experimental investigation of chemical and biological periodic phenomena is therefore of fundamental importance for biophysics, biochemistry, physiology, and for biology as a whole.

16.2. Belousov-Zhabotinsky Reactions

We shall begin our consideration of periodic processes with chemical auto-oscillatory phenomena.

In 1959 Belousov reported the discovery of periodic changes in the colour of a solution with a frequency of the order of 10^{-2} Hz in the course of the oxidation of citric acid by bromate with cerium ions used as the catalyst. The change of the colour was determined by the periodic conversions $Ce^{3+} \rightleftharpoons Ce^{4+}$. Belousov observed tens of periods. Later, detailed experimental and theoretical studies of these and related phenomena were carried out by Zhabotinsky and his coworkers (1964-1974).

The Belousov-Zhabotinsky reactions are oxidation-reduction reactions. Let us consider, as an example, the reaction in which the oxidant is bromate (the BrO_3^- ion), the reductant is malonic acid (MA, $H_2C(COOH)_2$), with cerium ions used as the catalyst. The oscillations in the colour determined by changes in the concentration of Ce^{4+} are relaxational in character (see Fig. 16.1*a*). Their period τ consists of two phases, τ_1 and τ_2—the increase and decrease of the concentration $[Ce^{4+}]$. The simplified scheme of the reaction involves two corresponding steps:

$$Ce^{3+} \xrightarrow{BrO_3^-,\ H^+} Ce^{4+} \quad \text{(I, oxidation)}$$

$$Ce^{4+} \xrightarrow{MA} Ce^{3+} \quad \text{(II, reduction)}$$

The products of the reduction of the bromate formed in step I brominate the malonic acid. The reaction yields bromomalonic acid (BMA), which is then decomposed to give Br^-. The bromide ion strongly inhibits the reaction.

The oscillatory process occurs as follows. Originally, the system contains a certain amount of Ce^{4+} ions. Step II gives rise to Br^- ions which inhibit reaction I and disappear at a certain rate from the system. The concentration of Ce^{4+} decreases as a result of reduction II and because of reaction I being inhibited by the bromide.

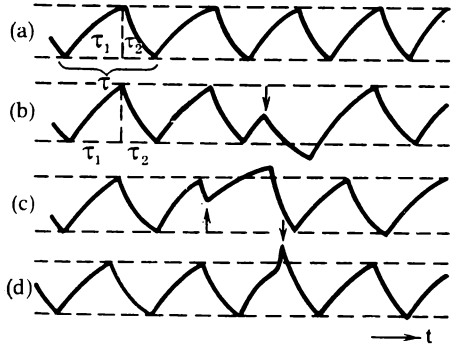


Fig. 16.1. The auto-oscillations of the concentration of Ce^{4+} (*a*). The discontinuities of the phases in the course of oscillations of Ce^{4+} caused by single additions of Br^- (*b*), Ag^+ (*c*), Ce^{4+} (*d*). The arrows indicate the moments of addition.

When $[Ce^{4+}]$ consumed in reaction II drops to the lower threshold value, the concentration of Br^- is also diminished. Reaction I starts again and the concentration of Ce^{4+} increases. When the upper threshold value of $[Ce^{4+}]$ is reached, $[Br^-]$ sharply increases, leading to the inhibition of reaction I, following which the cycle repeats. This scheme is supported by the following experimental facts:

1. Small amounts of Br^- introduced into the system in the phase of increase of $[Ce^{4+}]$ cause a phase shift in the chemical oscillation (Fig. 16.1*b*). The minimum amount of Br^- that causes a phase shift decreases by the end of the phase τ_1.

2. The Br^- ions added in the phase τ_2 lengthen it. The effect is the stronger, the greater is the amount added and the closer it is to the end of τ_2.

3. The addition of Ag^+, which binds Br^-, causes opposite events. The introduction of Ag^+ into the phase τ_2 leads to a phase shift (Fig. 16.1*c*), and the addition in the ascending phase increases τ_1.

4. If Br^- is continuously gradually added to the system, which is in the oscillatory regime, the oscillations are interrupted and the concentration $[Ce^{4+}]$ remains at the lower threshold. After the supply of Br^- is stopped the oscillations are resumed.

5. A similar result is obtained upon gradual addition of Ce^{4+}.

6. The addition of Ce^{4+} to the phase τ_1 causes a phase shift (Fig. 16.1*d*).

The general scheme of the reaction is shown in Fig. 16.2.

These experiments are highly efficient—the colour of the solution periodically changes from blue to pink and back again.

In auto-oscillatory reactions of this kind the reducing agent must be readily oxidized by the oxidized form of the catalyst and must not react directly with the bromate. It is also necessary that the reductant be readily brominated and the bromo derivatives be decomposed with the liberation of Br^-. A number of reactions involving various reducing agents have been studied.

Direct experiments have demonstrated that the reactions are quite homogeneous; the surface of the reaction vessel has no effect either on the course of a given periodic reaction or on the dependence of stability on reactant concentration.

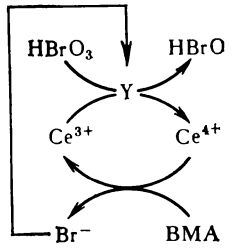


Fig. 16.2. The general scheme of a periodic chemical reaction.

The Belousov-Zhabotinsky reactions are complicated; it is difficult to take account of all their steps. The number of steps is presumably not less than 11. The problem can, however, be simplified and the oscillations in a homogeneous solution can be accounted for by considering three key substances: (1) $HBrO_2$ which plays the role of an intermediate switch; (2) Br^- which is a controlling intermediate substance; and (3) the catalyst Ce^{4+}. The following reactions prevail:

(a) at a sufficient amount of Br^-

$$BrO_3^- + Br^- + 2H^+ \xrightarrow{k_1} HBrO_2 + HBrO \qquad (I)$$

$$HBrO_2 + Br^- + H^+ \xrightarrow{k_2} 2HBrO \qquad (II)$$

(b) at a small amount of the Br^- left

$$BrO_3^- + HBrO_2 + H^+ \xrightarrow{k_3} 2BrO_2 + H_2O \qquad (III)$$

$$2BrO_2 + 2Ce^{3+} + 2H^+ \xrightarrow{k_4} 2HBrO_2 + 2Ce^{4+} \qquad (IV)$$

$$2HBrO_2 \xrightarrow{k_5} BrO_3^- + HBrO + H^+ \qquad (V)$$

The first step determines the rate. Hypobromous acid, HBrO, is rapidly consumed for the bromination of malonic acid. In steps (I) and (II) there is attained a quasi-stationary state (A):

$$k_2[HBrO_2]_A[Br^-][H^+] \approx k_1[BrO_3^-][H^+]^2[Br^-]$$

that is,

$$[HBrO_2]_A \approx (k_1/k_2)[BrO_3][H^+]$$

with $k_1/k_2 = 10^{-9}$.

The rate of the subsequent reactions is limited by step (III). Steps (III) and (IV) are jointly autocatalytic with respect to $HBrO_2$. The new quasi-stationary state (B) corresponds to the condition

$$k_5[HBrO_2]_B^2 \approx k_3[BrO_3^-][HBrO_2][H^+]$$

or

$$[HBrO_2]_B \approx (k_3/k_5)\cdot[BrO_3^-][H^+]$$

where $k_3/k_5 \approx 10^{-4}$ and $k_3 = 10^4\ M^2s^{-1}$.

Steps (II) and (III) show that Br^- and BrO_3^- compete for $HBrO_2$. The autocatalytic production of $HBrO_2$ is impossible until $k_2[Br^-] > > k_3[BrO_3^-]$. At the critical bromide concentration

$$[Br^-]_{cr} = \frac{k_3}{k_2}[BrO_3^-] \approx 5 \times 10^{-6}[BrO_3^-]$$

the reaction switches over from (I), (II) to (III), (IV). As the concentration $[HBrO_2]$ increases the concentration $[Br^-]$ decreases according to step (II) and passes through the critical level. On the other hand, the Ce^{4+} ions produced in step (IV) regenerate Br^-:

$$4Ce^{4+} + BrCH(COOH)_2 + H_2O + HBrO \xrightarrow{k_6} 2Br^- + 4Ce^{3+} + 3CO_2 + 6H^+ \quad (VI)$$

The concentration $[Br^-]$ again exceeds $[Br^-]_{cr}$ and the level of $[HBrO_2]_A$ is restored.

The mathematical model of the process (called the *Oregonator*) is as follows. We introduce the notations: $X = [HBrO_2]$, $Y = [Br^-]$, $Z = 2[Ce^{4+}]$, $A = [BrO_3^-]$, P and Q are the remaining products. The equations (I) through (VI) are rewritten as follows:

$$A + Y \xrightarrow{k_1} X, \quad X + Y \xrightarrow{k_2} P, \quad A + X \xrightarrow{k_{3,4}} 2X + Z,$$

$$2X \xrightarrow{k_5} Q, \quad Z \xrightarrow{k_6} \nu Y$$

Here ν is the appropriate stoichiometric coefficient. The rate constants evaluated from experiment are as follows: $k_1 = 1.34\ M^{-1}s^{-1}$, $k_2 = 1.6 \times 10^9\ M^{-1}s^{-1}$, $k_{3,4} = 8 \times 10^3\ M^{-2}s^{-1}$, $k_5 = 4 \times 10^7\ M^{-1}s^{-1}$. The values of ν and k_6 are the parameters. The following quantities are assumed to be constant: $A = 0.06$ M and $[H^+] = 0.8$ M. The non-

linear equations of the irreversible reactions (i.e., those far from equilibrium) are:

$$\begin{aligned}\dot{X} &= k_1AY - k_2XY + k_{3,4}AX - 2k_5X^2\\ \dot{Y} &= -k_1AY - k_2XY + \nu k_6Z\\ \dot{Z} &= k_{3,4}AX - k_6Z\end{aligned} \tag{16.1}$$

We now introduce the dimensionless variables

$$\begin{gathered}x = \frac{k_2}{k_1A}X, \quad y = \frac{k_2}{k_{3,4}A}Y, \quad z = \frac{k_2k_6}{k_1k_{3,4}A^2}Z\\ \tau = A\sqrt{k_1k_{3,4}}\,t, \quad q = 2\frac{k_1k_5}{k_2k_{3,4}}, \quad s = \sqrt{\frac{k_{3,4}}{k_1}}\\ w = \frac{k_6}{A\sqrt{k_1k_{3,4}}}\end{gathered} \tag{16.2}$$

So we obtain

$$\begin{aligned}\frac{dx}{d\tau} &= s(y - xy + x - qx^2)\\ \frac{dy}{d\tau} &= \frac{1}{s}(-y - xy + \nu z)\\ \frac{dz}{d\tau} &= w(x - z)\end{aligned} \tag{16.3}$$

From the above estimates it follows that $q = 8.4 \times 10^{-6}$, $s \approx 80$, and $w = 0.16k_6$. The system has two time scales—the evolution of x is determined by the large constant s and the evolution of y and z is determined by the small constants.

The steady-state solutions of Eqs. (16.3) are as follows:

$$z_0 = x_0, \quad y_0 = \frac{\nu x_0}{1 + x_0} = \frac{1}{2}[(1 + \nu) - qx_0] \tag{16.4}$$

The second equation in (16.4) is equivalent to that obtained from the first equation in (16.3):

$$qx_0^2 + [q - (1 - \nu)]x_0 - (1 + \nu) = 0$$

Linearization according to the known rule (page 521) yields:

$$x = x_0 + \alpha e^{\lambda\tau}, \quad y = y_0 + \beta e^{\lambda\tau}, \quad z = z_0 + \gamma e^{\lambda\tau}$$

This system is more complex than those considered earlier (the Lotka-Volterra model and the Brusselator)—in this case we have a third-order characteristic equation instead of the second-order one:

$$\lambda^3 - f\lambda^2 + g\lambda - h = 0 \tag{16.5}$$

where f, g, and h depend on x_0, y_0, and z_0 and also on the parameters s, w, ν, and q, the values of ν and w being varied. An auto-

oscillatory process, i.e., a limit cycle, arises if the point (x_0, y_0, z_0) is unstable. It can be shown that the region of such solutions does really exist if

$$0.50 < \nu < 2.41$$

Figure 16.3 shows a limit cycle in the log X, log Y plane.

The oscillatory regime can be controlled through external influences, such as a constant influx of Br^- or the action of ultraviolet light on the solution, which leads to the same results, since this light decomposes bromomalonic acid with the evolution of Br^-. The auto-oscillations are synchronized upon action of a periodic external force—rectangular light pulses with a frequency of recurrence close to the frequency of auto-oscillations. Of course, the process is stopped when the entire amount of malonic acid present is oxidized to CO_2 (reaction (VI)).

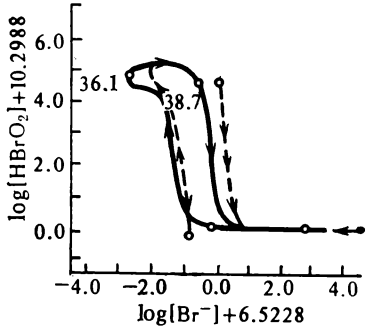


Fig. 16.3. The limit cycle for the Oregonator.

We have examined a point, concentrated system. If the system is a distributed system, i.e., if chemical reactions are accompanied by diffusion (see Sec. 15.4), wavelike processes can occur in it, whose space and time characteristics do not depend on the initial conditions—they are referred to as *autowave processes*. Such processes apparently play an important role in many biological phenomena—in morphogenesis (Sec. 17.7) and in the behaviour of excitable tissues (Sections 16.5 and 16.6), etc.

Zhabotinsky and Zaikin have observed and studied the autowave auto-oscillatory processes in the chemical system described above. For the system to be distributed, no convection must occur in the presence of diffusion. This is accomplished in thin tubes (a one-dimensional system) or in thin layers (a two-dimensional system). The catalyst used is a ferroin complex of iron instead of Ce^{3+}. The system is initiated by touching the surface of the solution with a needle moistened with a solution of $AgNO_3$ (the Ag^+ ion binds Br^-) or with a hot filament. There appear waves which run at a velocity of about 0.01 cm/s. There have been experimentally detected point sources of auto-generation—*leading centres* (LC) that arise in a homogeneous medium as a result of local fluctuations in concentration, i.e., special initial conditions. The concentration waves spread out from a leading centre at constant velocity. Figure 16.4 shows the

leading centres photographed in the system indicated at 30-second intervals.

A break in the wavefront may give rise to a *reverberator*—a spiral wave. Spiral waves are generated, for example, during the mo-

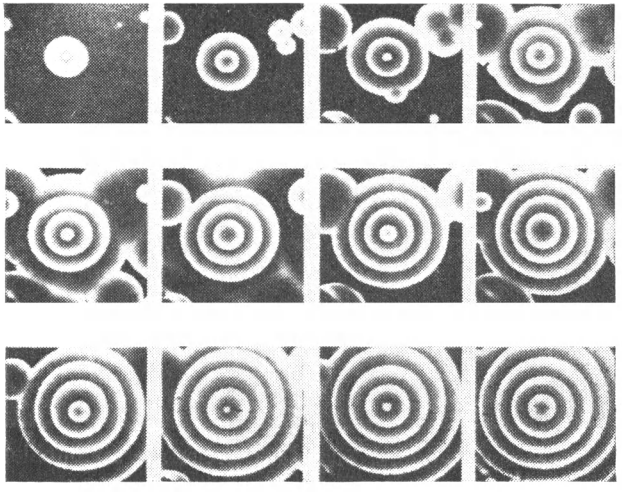


Fig. 16.4. The leading centres.

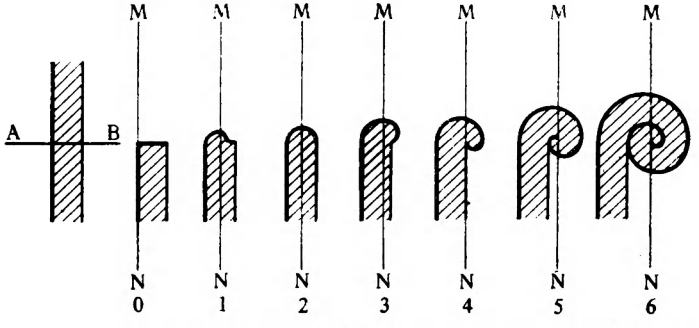


Fig. 16.5. Schematic representation of the formation of a reverberator from a break in the wavefront.

The solid line represents the zone of stimulation; *AB* is the line showing the break in the front; 0-6—successive moments of time; MN—the front line at zero time.

tion of a wave in a two-dimensional medium around an opening—a spiral wave is the involute of the opening.

The scheme of formation of a reverberator is given in Fig. 16.5. Zhabotinsky and Zaikin observed spiral waves initiated by a distur-

Fig. 16.6. Chemical reverberators. The black circles are air bubbles.

Fig. 16.7. The lichen *Parmelia centrifuga*.

bance in the system's homogeneity—a local decrease in acidity. Figure 16.6 shows a photograph of chemical spiral waves and Fig. 16.7 is a photograph of lichen. An external resemblance is obvious. Of course, the periodic growth of lichen is not the analogue of the behaviour of the chemical systems discussed above, but there is every ground for believing that chemical auto-oscillatory and autowave processes model important biological phenomena.

16.3. Auto-Oscillations in Glycolysis

The nonlinearities in the kinetics of biochemical, enzymatic processes manifest themselves, in a number of cases, in the periodicity observed. The origin of oscillations in biochemical systems is on the whole understandable—they are determined by autocatalysis and in this sense are modelled by periodic reactions in nonliving things.

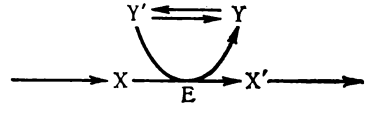


Fig. 16.8. A substrate-inhibited reaction.

We have already studied nonlinear biochemical processes—cooperative enzymatic reactions (Sec. 6.7). Chemical processes that obey the law of mass action are, as a rule, nonlinear. The interactions between the elements of a multienzyme system and the interactions of such systems with each other are non-linear. Typical nonlinear enzymatic processes are reactions inhibited by an excess of substrate or coenzyme. The reaction

$$X+Y' \xrightarrow{E} X'+Y$$

catalyzed by the enzyme E is inhibited by the active form of the coenzyme Y'. The scheme of the reaction is shown in Fig. 16.8. Under certain conditions the system is characterized by unstable steady states and may become auto-oscillatory (Selkov).

Periodic, auto-oscillatory phenomena are characteristic of the process of *glycolysis*—the anaerobic conversion of six-membered sugars to tricarboxylic acids leading to the synthesis of ATP (see Sec. 2.10). The simplified scheme of the process is presented in Fig. 16.9. In their experiments Chance and Hess and coworkers (1964) first observed damped concentration oscillations in glycolysis and then discovered undamped auto-oscillations. Figure 16.10 shows oscillations in the concentration of one of the products of glycolysis, $NADH_2$, which are almost sinusoidal in shape. As has been shown by Selkov, the kinetics of the process is determined on the whole by several "narrow sites" marked in Fig. 16.9 by the numbers *1-4*. Let v_0 be the

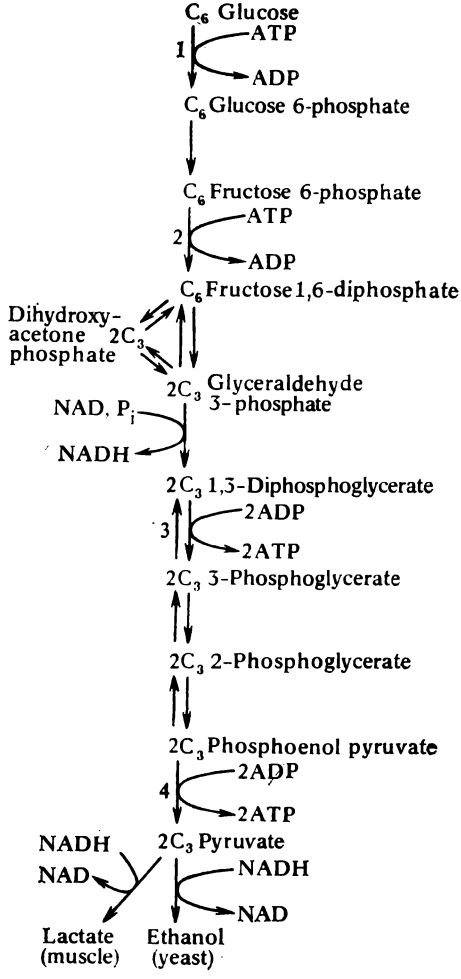


Fig. 16.9. A simplified scheme of the glycolysis:
C_6 and C_3 are six- and three-carbon compounds; *1-4*—reactions which determine the kinetics of the process as a whole.

rate of reaction *1* catalyzed by the enzyme hexokinase (HK). The reaction $C_6' \rightleftharpoons C_6''$ is fast and reversible, and therefore the concentrations of these hexoses may be assumed to be proportional to each other. The concentration X of any of the substances, C_6' and C_6'', increases at a rate v_0 and decreases due to reaction *2* catalyzed by phos-

phofructokinase (PFK). This is followed (up to reaction *3*) by several fast reversible reactions and, hence, the concentrations of the intermediate compounds are proportional to each other. Let us denote one of these concentrations by Y. Phosphofructokinase is activated by the products of its reaction, i.e., by compound Y. Therefore, the increase of Y due to reaction *2* depends both on X and on Y.

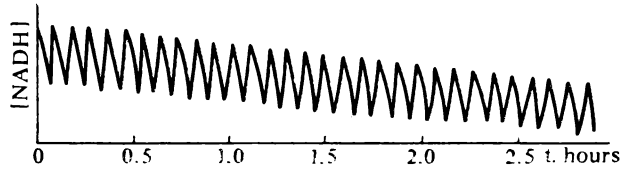


Fig. 16.10. The oscillations of the concentration of NADH in a yeast extract.

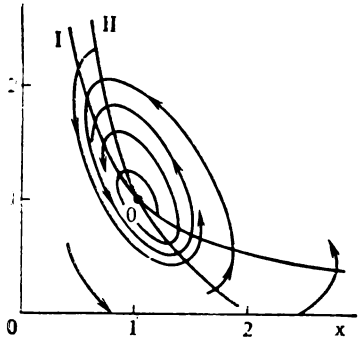


Fig. 16.11. The phase pattern of oscillations during glycolysis.

The rate of increase of Y is equal to the rate of decrease of X. In reaction *3* catalyzed by pyruvate kinase (PK) the concentration Y is reduced. We obtain the following kinetic equations:

$$\begin{aligned} \dot{X} &= v_0 - k_1 \frac{X}{K_X + X} \frac{Y}{K_Y + Y} \\ \dot{Y} &= k_1 \frac{X}{K_X + X} \frac{Y}{K_Y + Y} - k_2 \frac{Y}{K + Y} \end{aligned} \tag{16.6}$$

In the work of Selkov (1968), instead of $Y/(K_Y + Y)$ there appears the term $Y^2/(K_Y + Y)^2$, i.e., a higher order of activation of phosphofructokinase is introduced, this being substantiated by biochemical data. However, even the simple scheme (16.6) has a nonlinear character of interest to us. Moreover, the auto-oscillatory regime al-

so appears with further simplification if we assume that $K_X \gg X$ and $K_Y \gg Y$. We obtain the system

$$\begin{aligned} \dot{X} &= v_0 - k_1' XY \\ \dot{Y} &= k_1' XY - k_2' \frac{Y}{1+Y/K} \end{aligned} \tag{16.7}$$

which is similar to the Lotka system (see page 515). Here $k_1' = k_1 K_X K_Y$ and $k_2' = k_2/K$; v_0 is proportional to the content of hexokinase, k_1' to the content of phosphofructokinase, and k_2' to the content of pyruvate kinase.

The steady-state concentrations of X and Y are equal to

$$X_0 = \frac{k_2' - v_0/K}{k_1'}, \quad Y_0 = \frac{v_0}{k_2' - v_0/K}$$

We introduce the dimensionless variables

$$x = \frac{X}{X_0}, \quad y = \frac{Y}{Y_0}, \quad \tau = k_1' Y_0 t = \frac{k_1' v_0}{k_2' - v_0/K} t$$

Equations (16.7) take the form

$$\frac{dx}{d\tau} = 1 - xy, \quad \frac{dy}{d\tau} = qy\left(x - \frac{1+r}{1+ry}\right) \tag{16.8}$$

where

$$q = \frac{(k_2' - v_0/K)^2}{v_0 k_1'}, \quad r = \frac{Y_0}{K} = \frac{v_0/K}{k_2' - v_0/K}$$

The phase pattern of the system in the x, y-plane and the principal isoclines (see page 506) are shown in Fig. 16.11. The isocline of the vertical tangent lines (I) is the hyperbola $y = 1/x$ and that of the horizontal tangent lines (II) is the hyperbola $x = (1 + r)/(1 + ry)$ and the straight line $y = 0$. Linearizing Eq. (16.8), we determine the character of the singular points. The roots of the characteristic equation are given by

$$\lambda_{1,2} = -\frac{1}{2}\left(1 - \frac{qr}{1+r}\right) \pm \frac{1}{2}\left[\left(1 - \frac{qr}{1+r}\right)^2 - \frac{4q}{1+r}\right]^{1/2}$$

At $4q/(1 + r) > (1 - qr/(1 + r))^2$ the point 0 is a focus. If $qr/(1 + r) < 0$, the focus is stable; if $qr/(1 + r) > 0$, the focus is unstable and there appears a limit cycle. At $qr/(1 + r) = 1$ bifurcation occurs. Near this point (if $| qr/(1 + r) - 1 | \ll 1$) the frequency of oscillations is equal to the imaginary part of the root of the characteristic equation, i.e.,

$$\omega \approx \sqrt{\frac{q}{1+r}} \tag{16.9}$$

Thus, the system (16.7) proves to be auto-oscillatory under certain conditions. The increase of the parameter q contributes to the amplification of auto-oscillations, and the decrease of q to their damping. From the expression for q it follows that the decrease of k_1' and of the rate of influx v_0 must cause the self-excitation of oscillations. The same result is obtained when k_2' increases, i.e., by the addition of pyruvate kinase. This is in agreement with experiment.

16.4. Auto-Oscillations in Photosynthesis

Photosynthesis (see Chapter 14) undergoes periodic oscillations in intensity. Presumably, the periodicity determined by the endogenous rhythm is characteristic of the Calvin dark cycle (page 476), since the reactions of the light cycle proceed very rapidly, for fractions of a second. The rhythm associated with the dark cycle is characterized by a period which is substantially shorter than 24 hours but is synchronized with the periodicity of illumination.

The directly observable rhythm of photosynthesis must evidently be determined by periodic biochemical processes of the dark cycle. The theoretical model of oscillations in the dark reactions of photosynthesis has been constructed by Chernavsky and his coworkers (1958).

The Calvin cycle involves three-, five-, and six-carbon compounds, C_2, C_5, and C_6. The consideration of all the kinetic equations for the cycle is too complicated. It is reasonable to simplify the system as much as possible with the system retaining its principal qualitative features. The kinetics of the process is limited by the slow steps, and the fast reversible steps may be treated as equilibrium steps. Therefore, one should confine oneself to the consideration of the rate-limiting steps.

We shall consider only two kinetic equations, in which the variables are the concentrations C_3 and C_6. The remaining variables are expressed in terms of C_3 and C_6 algebraically. Since the Calvin cycle mainly involves bimolecular reactions, the kinetics is expressed by second-order equations:

$$\begin{aligned} \dot{C}_3 &= \alpha_1 C_3^2 + \alpha_2 C_3 C_6 + \alpha_3 C_6^2 \\ \dot{C}_6 &= \beta_1 C_3^2 + \beta_2 C_3 C_6 + \beta_3 C_6^2 \end{aligned} \tag{16.10}$$

with $\alpha_3 = 0$, since the concentrations of the trioses is not changed as a result of the interaction between the two hexoses. The increase of C_6 may occur due only to the combination of two molecules of C_3. Hence, $\beta_1 > 0$ and β_2, $\beta_3 < 0$. An increase in the content of trioses occurs in a more complex process involving both the trioses themselves and the hexoses C_6. We may put $\alpha_1 > 0$ and $\alpha_2 < 0$. We obtain

the equations

$$\begin{aligned} \dot{C}_3 &= \alpha_1 C_3^2 - \alpha_2 C_3 C_6 \\ \dot{C}_6 &= \beta_1 C_3^2 - \beta_2 C_3 C_6 - \beta_3 C_6^2 \end{aligned} \tag{16.11}$$

where all α_i and $\beta_i > 0$.

We find the isoclines and steady states corresponding to $\dot{C}_3 = \dot{C}_6 = 0$. The two isoclines are straight lines passing through the coordinate origin, their equations being $C_6 = a_1 C_3$ and $C_6 = a_2 C_3$, where a_1 and a_2 satisfy the conditions $\alpha_1 - a_1\alpha_2 = 0$ and $\beta_1 - \beta_2 a_2 - \beta_3 a_2^2 = 0$. There is only one steady state, $C_3^0 = C_6^0 = 0$ at $a_1 \neq a_2$ or an infinite set of such states at $a_1 = a_2$. In the first case no photosynthesis occurs, and the second corresponds to indifferently unstable states. No limit cycles and auto-oscillations take place. Hence, the system (16.11) is insufficient and the model requires completion. We introduce into the first equation an additional constant term, $\alpha_0 > 0$, which describes the influx of C_3 that arises in other processes different from the Calvin cycle (glycolysis, respiration). Such an influx serves as the initiator of the cycle. The necessity of the triggering mechanism is characteristic of autocatalysis. So,

$$\begin{aligned} \dot{C}_3 &= \alpha_1 C_3^2 - \alpha_2 C_3 C_6 + \alpha_0 \\ \dot{C}_6 &= \beta_1 C_3^2 - \beta_2 C_3 C_6 - \beta_3 \alpha_6^2 \end{aligned} \tag{16.12}$$

We now pass to the variables $X = C_3/C_3^0$ and $Y = C_6/C_6^0$, where C_3^0 and C_6^0 are the steady-state values of C_3 and C_6 obtained from the conditions $\dot{C}_3 = \dot{C}_6 = 0$. We have

$$\begin{aligned} \dot{X} &= \alpha_1' X^2 - \alpha_2' XY + \alpha_6' \\ \dot{Y} &= \beta_1' X^2 - \beta_2' XY - \beta_3' Y^2 \end{aligned} \tag{16.13}$$

where $\alpha_1' = \alpha_1 C_3^0$, $\alpha_2' = \alpha_2 C_6^0$, $\alpha_0' = \alpha_0/C_3^0$, $\beta_1' = \beta_1 (C_3^0)^2/C_6^0$, $\beta_2' = \beta_2 C_3^0$, $\beta_3' = \beta_3 C_6^0$. In the steady state $X = Y = 1$. Hence,

$$\begin{aligned} \alpha_1' - \alpha_2' + \alpha_0' &= 0 \\ \beta_1' - \beta_2' - \beta_3' &= 0 \end{aligned} \tag{16.14}$$

In the steady-state regime the coefficient β_1' is the quantity of hexose formed per unit time from the triose; β_2' is the amount of hexose removed from the Calvin cycle and forming starch; β_3' is the amount of hexose returning to the cycle. For one hexose molecule to be removed, six revolutions of the cycle are required, since only one molecule of CO_2 can be used in the cycle. Therefore, $\beta_2' = \beta_3'/6$. Thus,

$\beta_3' = 6\beta_1'/7$, $\beta_2' = \beta_1'/7$. On the other hand, $\alpha_0' \ll \alpha_1', \alpha_2'$. Introducing the dimensionless time $\tau = \alpha_1' t$, we have

$$\begin{aligned} \frac{dX}{d\tau} &= X^2 - \frac{\alpha_2'}{\alpha_1'} XY + \frac{\alpha_0'}{\alpha_1'} \\ \frac{dY}{d\tau} &= \frac{\beta_1'}{\alpha_1'} \left(X^2 - \frac{1}{7} Y^2 - \frac{6}{7} XY \right) \end{aligned} \tag{16.15}$$

We linearize the system by putting

$$X = 1 + x, \quad Y = 1 + y, \quad x, y \ll 1$$

Eliminating α_2' with the aid of the first relation in (16.14), we obtain from Eqs. (16.15)

$$\begin{aligned} \frac{dx}{d\tau} &= (1 - \gamma)\, x - (1 + \gamma)\, y \\ \frac{dy}{d\tau} &= \frac{8}{7} \varepsilon\, (x - y) \end{aligned} \tag{16.16}$$

where $\gamma = \alpha_0'/\alpha_1'$ and $\varepsilon = \beta_1'/\alpha_1'$. We look for the solution of the form

$$x = x_0 e^{\lambda\tau}, \quad y = y_0 e^{\lambda\tau}$$

We find the roots of the characteristic equation:

$$\lambda_{1,2} = \frac{1}{2} (1 - \gamma - 8/7\varepsilon) \pm [1/4\, (1 - \gamma - 8/7\varepsilon)^2 - 16/7\varepsilon\gamma]^{1/2} \tag{16.17}$$

The steady state $x = y = 0$ is stable if the real part of λ is negative, i.e., if

$$1 - \gamma - 8/7\varepsilon < 0$$

and unstable in the opposite case. The transition to the unstable regime occurs at $\varepsilon = \frac{7}{8} (1 - \gamma) \approx 0.87$ (since $\gamma \ll 1$). When α_1' increases, such a transition proves possible.

The imaginary part of (16.17) determines oscillations. Their dimensionless period is equal to $T = 2\pi/\mathrm{Im}\lambda$, where the imaginary part of λ is $\mathrm{Im}\,\lambda = \pm [\frac{16}{7} \varepsilon\gamma - (\mathrm{Re}\lambda)^2]^{1/2}$. Upon transition to the unstable regime, when $\mathrm{Re}\,\lambda = 0$, we have

$$T = \frac{2\pi}{\sqrt{\frac{16}{7} \varepsilon\gamma}}$$

The behaviour of the system is studied by means of the ordinary method of phase mapping. To specify we set $\gamma \approx \frac{1}{20}$ and $\varepsilon \approx \frac{2}{3}$.

Then, the instability condition is satisfied, since $\frac{19}{20} > \frac{16}{21}$. The integral curves are described by the equation

$$\frac{dy}{dx} = \frac{2}{3} \frac{x^2 - \frac{1}{7} y^2 - \frac{6}{7} xy}{x^2 - \frac{21}{20} xy + \frac{1}{20}} \tag{16.18}$$

Figure 16.12 shows the phase pattern of the system (16.16) at the indicated values of γ and ε. The system is auto-oscillatory; C_3 and

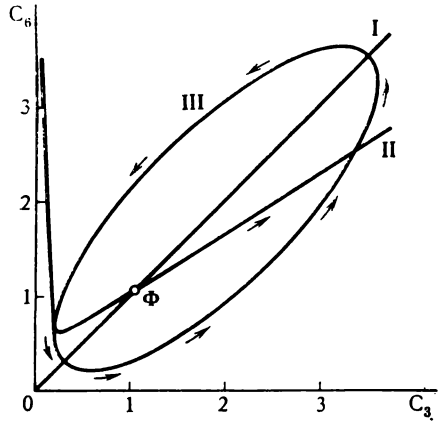


Fig. 16.12. The phase pattern for the photosynthesis model:
I and *II*—lines of parallel and perpendicular isoclines; *III*—limit cycle. Φ ($x = y = 1$)—stable focus.

C_6 change periodically with the same phase. A numerical or graphical determination of the period of oscillations of C_3 and C_6 indicates that this period is about two orders of magnitude greater than the characteristic time $1/\alpha_1'$. This time is estimated as 5-10 min, and the period T is of the order of several hours. The oscillations are anharmonic: the system resides most of time in the region of small concentrations of C_3 and C_6.

Computer calculations show that the system is synchronized by an external periodic influence (by light).

16.5. Nonlinear Dynamics of Membranes

As we have seen, the nonlinear properties of excitable membranes are distinctly manifested in the generation and propagation of a nervous impulse (Chapter 11). Let us consider periodic changes in the state of membranes established in a number of experiments. For instance, there were observed oscillations of electrical potential

in very thin polyelectrolyte bilayer membranes. The bilayer consisted of a polyacid (*a*) and a polybase (*b*). Thus, it had three zones—a negatively charged zone *a*, a neutral zone, and a positively charged zone *b* (Fig. 16.13). The membrane was placed in a 0.15M solution of NaCl. When a negative potential was applied, periodic impulses (spikes) were observed on the side of the polybase and at a certain critical value of current there were observed undamped oscillations, which persisted for hours. The current through the membrane consists of the movement of cations through the zone *a* and of anions through the zone *b*. As a result, NaCl is accumulated in the central neutral zone. The increase of the osmotic pressure leads to the appearance of a flux of solvent into the membrane and to an increase in the hydrostatic pressure in it. At the same time, the increase of the salt concentration causes the polyelectrolyte molecules to compress, which also increases the pressure. When this increase exceeds the osmotic pressure, the solvent flux changes its sign, and the salt concentration inside the membrane increases even more. A concentration gradient arises, the salt leaves the membrane and flows out after the membrane attains a maximum contraction. This is followed by relaxation, the membrane returns to the initial state, and the process starts again.

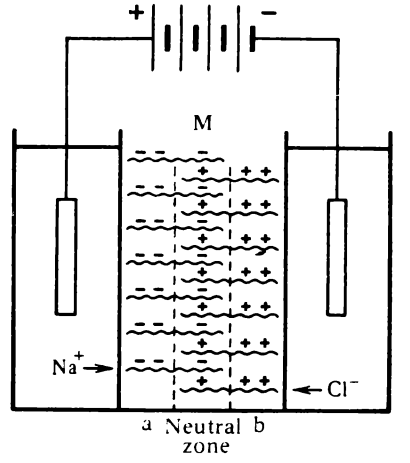


Fig. 16.13. A diagram for a three-layer polyelectrolyte membrane.

The *Teorell membrane oscillator* (1955) consists of two cells filled with an electrolyte of different concentrations and separated by a porous-glass membrane containing bound negative charges. When a direct current is passed, a potential difference is created on the two sides of the membrane and a shift in the liquid levels occurs in the cells. At high constant currents periodic oscillations in the liquid levels and in the membrane potential occur in the system. Because of the presence of negative charges in the membrane the electrolyte in its pores carries mainly cations. The external field causes the movement of the cations and the solvent. The concentration distribution in the membrane pores depends on the amount of the flowing liquid and its flow rate. In turn these factors govern the resistance of the membrane. The forces that act on the system depend on the state of

the system. There arise auto-oscillations, the theory of which has been developed by Teorell (1964), Markin (1971), and others.

In biological membranes containing enzymes, which have cooperative properties, the periodic changes in the state of the membrane may be determined by the nonlinear coupling between transport of matter and chemical reactions. We shall consider a simple example. Suppose that the enzyme catalyzes the irreversible degradation of substrate S. The chemical flux (see page 321) is given by

$$J_{\text{chem}} = v = -\frac{d_{\text{chem}}S}{dt} = kES \tag{16.19}$$

where S is the substrate concentration; E that of the enzyme; k is the rate constant. Suppose that there is a critical substrate concentration

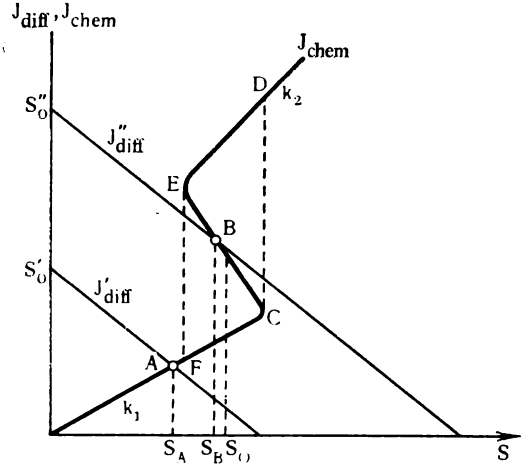


Fig. 16.14. Oscillations of the concentration in a coupled chemiosmotic system.

S_{cr} at which the enzyme undergoes cooperative conformational changes that lead to an increase in its activity. The rate constant k increases from k_1 to k_2. The dependence $J_{\text{chem}}(S)$ is presented in Fig. 16.14. Because of the cooperativity J_{chem} undergoes a transition similar to a phase transition (cf. page 527). The substrate enters the system by way of diffusion through the membrane at the rate

$$J_{\text{diff}} = \frac{d_{\text{diff}}S}{dt} = a\,(S_0 - S) \tag{16.20}$$

where S_0 is the concentration S in the source which is practically constant. At $J_{\text{chem}} = J_{\text{diff}}$ the rate of change of the concentration $\dot{S} = -\dot{S}_{\text{chem}} + \dot{S}_{\text{diff}} = 0$—the system is in a steady state. Figure 16.14 shows the difference between two steady states correspond-

ing to $S_0 = S_0'$ and $S_0 = S_0''$. The steady state corresponding to the point of intersection A of the curve for J_{chem} with the straight line for J'_{diff} is stable, since with a slight positive change of the substrate concentration $J_{\text{chem}} > J'_{\text{diff}}$ and the substrate concentration drops until the system returns to point A. With a negative fluctuation in S the flux $J'_{\text{diff}} > J_{\text{chem}}$ and S will increase until the system again returns to point A. On the contrary, if the point of intersection B with J_{chem} and J''_{diff} lies in the region of the curve for J_{chem} with a negative slope, the steady state B is unstable. To the positive fluctuation of S there corresponds $J''_{\text{diff}} > J_{\text{chem}}$ and S increases up to the point C. At this point the enzyme undergoes a conformational change, the system passes to point D, at which $J_{\text{chem}} > J''_{\text{diff}}$. The substrate concentration S falls down to point E at which the enzyme undergoes a reverse transition to the initial form, and the system moves to point F. Here S increases again, since $J''_{\text{diff}} > J_{\text{chem}}$ and the system again reaches the point C. Thus, the system evolves to the hysteresis loop $CDEF$ to which there corresponds the limit cycle on the phase pattern. Numerous examples could be cited for such behaviour of systems in which chemical reactions are combined with diffusion.

Of special interest in biology are excitable membranes. We have seen that the mathematical model of the nerve tissue is expressed, in particular, by the *Hodgkin-Huxley equations* (page 391). These nonlinear equations are of the fourth order. For the purposes of a qualitative study, we can, however, reduce the order of nonlinearity from four to two by eliminating differential equations for fast components (see page 514). One of the possible simplified second-order models has been constructed and studied by Krinsky, Ivanitsky and their coworkers.

In the Hodgkin-Huxley equations the slow variables are n and h ($\tau_m \sim 10^{-4}$ s, $\tau_n \sim \tau_h \sim 2 \times 10^{-3}$ s). We introduce the functions $\bar{m}(\varphi)$, $\bar{n}(\varphi)$, and $\bar{h}(\varphi)$, which express the steady-state values of the corresponding variables. As experiment shows, $\bar{n}(\varphi)$ and $\bar{h}(\varphi)$ are connected by the approximate relation

$$\bar{n}(\varphi) + \bar{h}(\varphi) \approx q = \text{const} \ (\approx 0.85) \tag{16.21}$$

The Hodgkin-Huxley equations can be simplified by replacing m by $\bar{m}(\varphi)$ and h by $q-n$. Then, the basic equations (11.5) through (11.9)

$$\begin{aligned} I = C\frac{\partial \varphi}{\partial t} + (\varphi - \varphi_{\text{K}})\,\bar{g}_{\text{K}} n^4 + (\varphi - \varphi_{\text{Na}})\,\bar{g}_{\text{Na}} m^3 h + (\varphi - \varphi_{\text{Y}})\,\bar{g}_{\text{Y}} \\ \dot{m} = [\bar{m}(\varphi) - m]/\tau_m(\varphi) \\ \dot{n} = [\bar{n}(\varphi) - n]/\tau_n(\varphi) \\ \dot{h} = [\bar{h}(\varphi) - h]/\tau_h(\varphi) \end{aligned} \tag{16.22}$$

yield the system

$$I = \alpha C \frac{\partial \varphi}{\partial t} + (\varphi - \varphi_K)\,\bar{g}_K n^4 + (\varphi - \varphi_{Na})\,\bar{g}_{Na}\bar{m}^3(\varphi)\,(q - n) + \\ + (\varphi - \varphi_Y)\,\bar{g}_Y \quad (16.23)$$

$$\dot{n} = [\bar{n}(\varphi) - n]/\tau_n(\varphi)$$

The parameter α is found to be equal to 3. Equations (16.23) give results which differ little from those obtained through the use of the initial model.

A qualitative study of the model with the aid of the phase pattern is carried out by plotting the isoclines $\dot{n} = 0$ and $\dot{\varphi} = 0$, i.e.,

$$n = \bar{n}(\varphi) \quad (16.24)$$

and

$$\bar{g}_K n^4 (\varphi - \varphi_K) + \bar{g}_{Na}(q - n)\,\bar{m}^3(\varphi)\,(\varphi - \varphi_{Na}) + \\ + \bar{g}_Y (\varphi - \varphi_Y) - I = 0 \quad (16.25)$$

The latter isocline has an N-shape in the n, φ-plane, which provides the generation of the impulse. The singular point at the intersec-

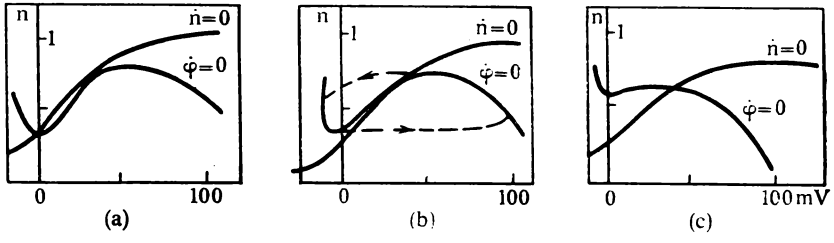


Fig. 16.15. The zero isoclines of the system:
(*a*) rest; (*b*) spontaneous activity; (*c*) stable depolarization

tion of the isoclines is stable (Fig. 16.15*a*), which corresponds to the resting state of the membrane. The spontaneous activity can be induced by an increase in the conductance $\bar{g}_K$ or by a decrease in $\bar{g}_{Na}$. It appears when the point of rest loses stability, this being followed by the emergence of a limit cycle and auto-oscillations (Fig. 16.15*b*). The singular points that lie on the middle branch of the $\dot{\varphi} = 0$ isocline are unstable. Accordingly, the further increase of the Na-current (for example, due to the shift of $m(\varphi)$ by 20 mV to the left) leads to the disappearance of the stable depolarization of the membrane (Fig. 16.15*c*). Repeated periodic responses arise at such a value of

the external current I in (16.25) at which the singular point is in the middle region of $\dot{\varphi} = 0$. With the aid of a qualitative analysis of the phase pattern one can find conditions for the triggered behaviour of the membrane. Under certain experimental conditions, say, on intracellular perfusion of CsF, the squid axon behaves as a trigger with two stable states—the stimulation by a current directed alternatively outwards and inwards leads to the flip from one stable state to another.

What has been discussed above shows that there is ground for treating changes in the state of an excitable membrane as phase transitions (see Sec. 15.5).

16.6. Autowave Processes in the Cardiac Muscle

As we have seen (Chapter 12), the muscle tissue is an *excitable medium*—the excitation imposed by a network of nerve fibres induces mechanochemical processes. The processes taking place in excitable media are therefore of greatest importance for physiology, in particular for the physiology of the cardiac muscle.

The concept of the excitable medium as a medium in which the propagation of the nerve impulse occurs without damping at the expense of the energy stored in cells was introduced by Wiener and Rosenblueth (1946). An impulse is generated by each point of the medium. An example of a one-dimensional excitable medium is a nerve fibre. Each point of the excitable medium may be in one of the three states: *rest*, *excitation*, and *refractory state* (see page 381). If a stimulus is applied to a small portion of the fibre in the resting state, it will become excited, the excited state persisting during the time τ, and then pass over to a refractory state of duration τ_R. In the refractory period τ_R the fibre section no longer responds to the exciting signal and, hence, none of the fibre portions can be excited with a frequency higher than τ_R^{-1}.

In excitable media there are realized auto-oscillatory processes described in Sec. 16.2. Wave sources may also arise in such media in which none of the elements is capable of being excited spontaneously, i.e., is not auto-oscillatory. The formation of such sources is associated with the special properties of the medium and the excitation wave.

In considering the Belousov-Zhabotinsky reactions we encountered two types of wave source—leading centres (echo) which generate concentric waves and reverberators, rotating spiral waves (see Figs. 16.4 and 16.6). The formation and subsequent multiplication of reverberators play a very important role in the mechanism of *fibrillation*—a very dangerous cardiac arrhythmia (Krinsky and Ivanitsky). Arrhythmia is a disturbance of the synchronization of wave propagation in the cardiac muscle—instead of synchronous contractions

there arises irregular disordered fluttering and the organism is rapidly destroyed if no defibrillation is made.

Spiral waves in a homogeneous two-dimensional system appear in the presence of an opening, whose perimeter exceeds the wavelength $\lambda = v\tau_R$ (v is the velocity of wave propagation). The circulation of the excitation wave (a reverberator) is also possible in a continuous medium with nonuniform refractory periods. If an impulse is sent to a certain ring-shaped region of the medium, it will not circulate, since the excitation waves propagating in a clockwise and a counterclockwise direction will meet and cancel each other. But if the refractory period is not the same everywhere, a circulation will arise due to the transformation of the rhythm.

Figure 16.16 shows an inhomogeneous medium consisting of two regions with different refractory periods, τ_{R1} and τ_{R2}. The interval

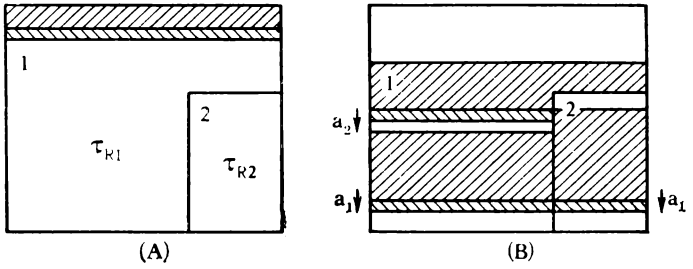


Fig. 16.16. Transformation of the rhythm in a two-dimensional inhomogeneous medium:
(*A*) distribution of refractory period ($\tau_{R1} > \tau_{R2}$); (*B*) break in the wavefront a_2.

Δt between two successive waves a_1 and a_2 is such that the wave a_2 cannot excite the region with $\tau_{R2} > \tau_{R1}$. In such a case the wavefront a_2 is disrupted. The refractory "tail" has a larger length in region *2* than in region *1*. The transformation of the rhythm consists in the cancellation of the second impulse.

A spiral wave arises from the contact of the wavefront with the tissue in the resting state. Such a situation is created, for example, by a counter wave b (Fig. 16.17*a*). The waves a_1 and b cancel each other out and the wave a_2 travels to the right half-plane in the form of a semi-circle (Fig. 16.17*b*) and then, bypassing its "refractory tail", runs to the left half-plane. The path is closed and there is formed a reverberator (Fig. 16.17*c*).

In a homogeneous medium, the period of waves originated from a reverberator is equal to τ_R, and the reverberator exists for an indefinitely long period of time. If it arises at the boundary between the

two regions with different τ_R, its lifetime is finite and the shorter, the greater is the inhomogeneity.

Reverberators may also serve as sources of high-frequency waves which cause the appearance of new reverberators. In other words, the multiplication of reverberators is possible. In an inhomogeneous medium, reverberators send waves of different frequencies and do not synchronize. If new wave sources appear more frequently than the old ones disappear, the phenomenon observed is complicated and similar to fibrillation.

The other wave source is *leading centres* or an *echo*. If several impulses are transmitted to a fibre that is inhomogeneous with respect to

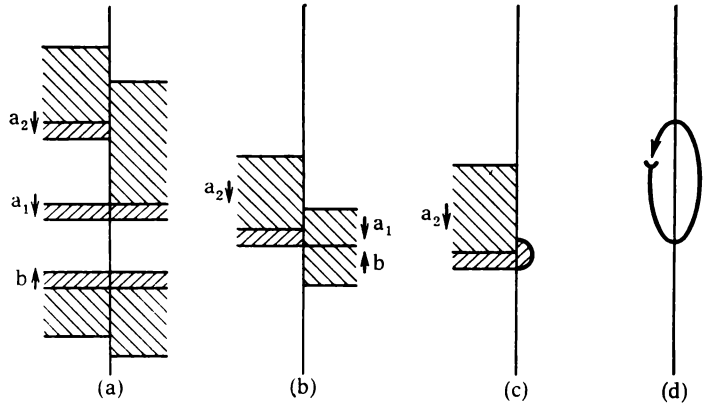


Fig. 16.17. The generation of a reverberator in an inhomogeneous medium: (*a*), (*b*), (c) successive steps; (*d*) reverberator.

refractory period, new impulses appear on the inhomogeneities which are running along the fibre and it is only some time later the resting state sets in. The impulses that arise are similar to those reflected from an inhomogeneity, which is regarded as a source of echo. It has been shown that an echo may appear in media for which $\tau/\tau_R > > 0.5$. The reverberator is a source of impulses capable of existence in a medium at $\tau/\tau_R > 0.5$ as well. The paths along which excitation passes from a region of lower τ_R to a region of higher τ_R and back are spatially separated but come closer together with increasing τ/τ_R and at $\tau/\tau_R = 0.5$ they converge. An echo is therefore a reverberator with a limiting, zero length. The echo and the reverberator are similar to each other as sources of impulses and may be responsible for fibrillation in inhomogeneous excitable media.

In electrophysiology, the state of the cardiac tissue is characterized by a so-called *vulnerable zone.* Vulnerability means the appearance of more than two responses when two impulses are conducted to a strip of myocardium. The width of the zone is the range of intervals between the impulses, Δt, at which vulnerability arises. Krinsky has shown that vulnerability is determined by the parameter $\tau' = \tau/\tau_R$; the zone width is the smaller, the lower is τ'.

In an inhomogeneous excitable medium, the wave source formed, whether an echo or a reverberator, may serve as a stimulus for starting new wave sources. If the parameters of the medium are such that the rate of "multiplication" of sources is not lower than the rate of their "death", fibrillation is very likely to occur. Upon interaction of several sources the duration of fibrillation increases rapidly with increasing τ'. The minimum number of sources capable of inducing fibrillation of a given duration is reduced with increasing τ'. Experiment shows that fibrillation on a strip of the myocardium tissue may be prolonged if the strip size is sufficiently large, i.e., if there is a certain *critical mass* for fibrillation. The number of sources that provides fibrillation is the analogue of the critical mass.

Thus, the dimensionless parameter τ' is an important characteristic of an excitable medium which determines the fibrillation regime and its duration.

As far as the cardiac muscle is concerned, it is necessary to find the analogue of the quantity τ expressing the duration of the excited state which can be measured by electrophysiological methods. The quantity introduced is the time of appearance of a response to the impulse applied, i.e., the *latent period* θ. It is known that the lower the amplitude of the impulse, the more delayed is the response. The quantity θ depends on time t that has elapsed after the last excitation of the cell by the time of arrival of the stimulus. There is a maximum delay θ_{max}, i.e., a response does not appear at the subthreshold amplitude of the impulse $A < A_{min}$ and at $t < t_{min}$. Thus, the function $\theta(t)$ describes the propagation of excitation. A theoretical analysis shows that to each value of θ there correspond two velocities of propagation of excitation. The higher velocity characterizes the stable process of "burning", whereas the lower velocity characterizes the unstable process ("smouldering") which is not observed on periodic conduction of the impulse under the steady-state regime. As the frequency of travel propagation of impulses increases the lower velocity increases and the higher velocity decreases. As the frequency increases further the two velocities coincide, the regime becomes nonstationary, the periodicity is disturbed and the impulses are cancelled out.

The formation of reverberators is facilitated when the parameter θ_{max}/τ_R increases. If this parameter is greater than 0.5, echo sources may be originated.

Thus, the vulnerability of the cardiac muscle is determined

by the quantity θ/τ_R. The antiarrhythmic substances used in medicine (called defibrillators) affect this parameter. An important practical result of theoretical investigations is that there has been found physical criteria for antiarrhythmogenic action that can be examined on tissue models by determining their effect on θ/τ_R. The factors influencing fibrillation are studied on the same basis. Indeed, it has proved possible to demonstrate experimentally that substances that lower θ/τ_R exert an antiarrhythmic effect and substances that increase θ/τ_R contribute to the appearance of cardiac arrhythmia, just as in models of excitable media a decrease in θ/τ_R hinders the formation of wave sources—an echo and a reverberator.

A detailed quantitative treatment of these phenomena has been carried out with the aid of a model of excitation propagation through a system of electrically connected cells. Analysis of the behaviour of the model is based on the method of phase patterns.

The theory of fibrillation of the cardiac muscle based on the general propositions of the physics of nonlinear oscillatory systems is an excellent example of a biophysical investigation culminating in important practical applications. The tasks of pathophysiology are distinctly formulated here as physical problems. We see that the physical approach to physiological problems provides rigour and generality of analysis and a sound substantiation of the inferences made.

Chapter 17 Problems of Biological Evolution and Development

17.1. The Origin of Life

We now turn our attention to the most general problem of biology and biophysics—the problems of development. As pointed out earlier, the basic specific features of living organisms are determined by their historical character—each organism develops ontogenetically and carries the memory of the phylogenetic, evolutionary development. Both ontogenesis and phylogenesis proceed in the direction of increasing complexity (see, however, Sec. 17.10) and are processes involving the generation and storage of new information and also, as we shall see later on, the increase of the information value (Sec. 17.9). Prebiotic evolution, which led to the formation of biological molecules, and biological evolution must be looked upon as part of the evolution of the Universe.

The first, fundamental problem which we encounter in connection with the problem of development is the problem of the origin of life. More than a hundred years ago (1863) Darwin wrote to Hooker: "It will be some time before we see 'slime, protoplasm, etc.', generating a new animal. But I have long regretted that I truckled to public opinion, and used the Pentateuchal term of creation, by which I really meant 'appeared' by some wholly unknown process. It is mere rubbish, thinking at present of the origin of life; one might as well think of the origin of matter."

Later (1871) Darwin did not consider it to be "mere rubbish" to speak of the origin of life and gave a clear-cut answer to the possibility of the origin of life in our time:

"It is often said that the conditions for the first production of a living organism are now present, which could ever have been present. But if (and oh! what a big if!) we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a proteine compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed."

The modern conceptions of the origin of life are intimately connected with the results obtained by the study of geological evolution. Oparin (1924) was the first to enunciate the theory of abiogenic origin of life. He presumed that the origin of life had been preceded by chemical evolution. Later, similar views were advanced by Haldane (1928). Today these ideas are widely accepted.

The primitive atmosphere of the Earth was a reducing atmosphere. Hydrogen is the main constituent of the Sun. The composition of the Sun is as follows: hydrogen 87%, helium 12.9%, oxygen 0.025%, nitrogen 0.02%, and carbon 0.01%. For comparison, let us cite the composition of the atmosphere of Jupiter: hydrogen 60%, helium 36%, methane about 1%, and ammonia about 0.05%. The ancient meteorites contain metals in the reduced form. As estimated by Miller and Urey, the partial pressure of H_2 in the primeval atmosphere of the Earth was 1.5×10^{-3} atm, the pressure of methane, CH_4, formed as a result of the reduction of carbon was 4×10^{-3} atm. The other components of the atmosphere were ammonia and water. These substances and also formaldehyde, H_2CO, are detected in the cosmic space.

The present-day atmosphere of the Earth contains a large amount of oxygen. No other planet of the Solar system contains such an amount of oxygen. Oxygen could have evolved as a result of the photodissociation of water under the influence of ultraviolet radiation of the Sun and as a result of photosynthesis. Bütner (1961) has shown that photodissociation must have given rise to appreciable amounts of oxygen. However, the primeval atmospheric oxygen was consumed for the oxidation of metals. It is commonly maintained that atmospheric oxygen is mainly of the biogenic, photosynthetic origin. The exact balance, however, has not yet been calculated.

Thus, photosynthesizing organisms referred to as *autotrophs* must have originated on the Earth during the period when its atmosphere was reducing. Moreover, if cells had not reached a highly organized and protected state, the atmospheric oxygen would not have made possible the development of life; it would have oxidized chemical compounds capable of giving rise to biologically functional molecules. The existence in our time of anaerobic bacteria and anaerobic glycolysis is evidence that life originated in the reducing atmosphere. Thus, the very existence of life itself points to the biogenic origin of the present-day oxidative atmosphere of the Earth. Geology and biology are inextricably connected.

Organic compounds must have been formed under reducing conditions in the presence of energy sources. Such sources were the ultraviolet radiation of the Sun, the radioactive radiation of the Earth (primarily, the β-decay of ^{40}K), and also electrical discharges in the atmosphere and the heat of volcanoes. The estimates of the amount of energy supplied to the Earth by all these sources are presented in Table 17.1.

Table 17.1. The Possible Sources of Energy for the Primary Chemical Evolution (in kJ per year per m^2)

Total solar radiation	10,900,000
Ultraviolet radiation, 300-250 nm	119,000
Ditto, 250-200 nm	22,000
Ditto, 200-150 nm	1,650
Ditto, <150 nm	71
Electrical discharges	
corona	126
lightning	42
Natural radioactivity down to a depth of 1 km, 4×10^9 years ago	117
Shock waves and pressure waves in the atmosphere	46
Solar wind	8
Volcanic heat	6
Cosmic rays	0.06

The possibility of monomeric organic compounds being synthesized under the conditions that simulate the conditions on the Earth with a reducing atmosphere has been proved by diverse experiments. Miller (1955) passed silent electrical discharges through mixtures of methane, ammonia, hydrogen, and water vapour. This gave rise to a mixture of a number of racemic amino acids and carboxylic acids, aldehydes and hydrogen cyanide, HCN. Terenin (1958) produced organic compounds by irradiating a mixture of CH_4, NH_3, H_2, H_2O with short-wavelength ultraviolet light. Amino acids are also formed in thermal synthesis, when such a mixture is passed through a hot tube at a temperature of the order of 1000°C. Amino acids are produced by bombarding a mixture of CH_4, NH_3, and H_2O in the solid state by electrons having an energy of 5 MeV. Under these conditions there is also formed adenine. Adenine and guanine are produced from HCN upon ultraviolet irradiation. Formaldehyde gives sugars, including ribose and deoxyribose. Nucleoside triphosphates are synthesized when a dilute solution of adenine, ribose, and phosphoric acid is exposed to ultraviolet radiation. It has been proved on the whole that the monomers of informational macromolecules of proteins and nucleic acids could have been synthesized in the primeval reducing atmosphere of the Earth. The source of purines and pyrimidines and also of porphyrins is HCN (cf. page 40). Amino acids and other organic compounds have been detected in ancient rocks and sediments and even in meteorites. No such substances are present on the Moon.

Mukhin presumed that the energy source required for the synthesis of monomeric organic compounds are underwater volcanoes. According to Table 17.1, volcanoes give only a small fraction of energy used for the synthesis of primary organics if it is calculated for the entire surface of the Earth. But if it is assumed that the synthe-

sis of an organic compound is determined by local events, the role of volcanism may prove rather significant. It should be kept in mind that the number of underwater volcanoes is much greater than that of land volcanoes. Mukhin has established the presence of HCN and organics in the matter ejected from volcanoes and demonstrated the formation of these compounds in model experiments (1974).

There is every ground for believing that the first stage of prebiotic evolution was the synthesis of monomeric organic compounds under far-from-equilibrium conditions. The next stage must have involved the polymerization and polycondensation of monomers, the formation of polypeptide and polynucleotide chains. A number of problems arise here.

First, a chemical system capable of evolution, i.e., of orderly development, must be autocatalytic and possess the ability to self-reproduce. Second, not any macromolecules can be the precursors of living systems, but only those which exhibit the required informational properties. Even at this stage of prebiotic evolution there must have existed selection mechanisms. As we shall see, Darwinian evolution must have been realized even at the prebiotic stage of development. The general problem is creation of information, the emergence of order from chaos.

In his works Fox (1966-1968) has experimentally studied self-organization in polypeptide chains formed by the thermal polycondensation of equimolar mixtures of 18 common amino acids (except Asp and Glu). This gives rise to chains of nonuniform composition, synthetic polypeptides, which were called *proteinoids* by Fox. These compounds display catalytic activity similar to enzyme activity; they have been used to carry out the reactions of hydrolysis, decarboxylation, amination, and deamination. No polymerase activity has, however, been detected in proteinoids.

Proteinoids are capable of forming *microspheres* if a suitable medium is chosen. There arises a *compartmentation* of the proteinoid system separated from the solution by a membrane-like shell. Fox maintains that such microspheres are the prebiological models of the cell. Oparin had earlier presumed that the compartmentation of primary polymers consisted of the formation of *coacervates*. The emergence of coacervate droplets is a phenomenon known in colloid chemistry; it means the separation of the solution from the solute. In this sense the Fox microspheres are similar to the Oparin coacervates. Anyway, one of the early stages of prebiotic evolution must have involved the appearance of compartmentation.

After the stages of formation of monomers and polymers, after compartmentation or simultaneously with it, there occurred, as one might think, a catalytic interaction between polynucleotide and polypeptide systems. The switch to biological evolution is probably associated with the origin of a genetic code.

The experimental abiogenic development of a living system is impossible so far. The dream of alchemists—the creation of the homunculus, a synthetic human being in the flask—is still far from being realized. But today science is capable of constructing informative physical models of prebiotic and biological evolution.

17.2. Modelling of Prebiotic Evolution

We shall here expound the *theory of self-organization of macromolecules* proposed by Eigen (1971).

Self-organization and *selection* are possible if the abiogenic molecular system is characterized by metabolism, self-reproduction, and mutability. These are the three prerequisites. Metabolism implies that the system is an open system in which polymerization and degradation of polymers take place. Since the system is far from equilibrium, these two processes are not interconnected by the condition of microscopic reversibility. To maintain metabolism, there is required a supply of a substance having an excess free energy—in the case of nucleic acids such substances are nucleoside triphosphates. Self-reproduction—the template replication of a polymer—implies an autocatalytic process. As has been shown in Chapters 15 and 16, autocatalysis can ensure self-organization. Finally, mutagenesis is necessary for creating new information.

Such a system may be called a Darwinian system if certain restricting conditions are imposed on it, say, the condition of constant organization, i.e., a constant overall concentration of energized monomers and a constant overall concentration of all types of polymers. For this constancy to be ensured, appropriate fluxes of monomers and polymers must be chosen.

An objection was raised against Darwin's theory of evolution which seemed convincing: natural selection means the survival of the fittest. But the criterion of fitness, adaptation, is survival. Hence, the Darwin theory is a vicious circle, a tautology, meaning, as it were, the survival of those who survive.

In fact, the criteria of adaptation which are satisfied by the highest survival of progeny are quite objective, being determined by the environmental conditions. It is the role of these conditions that is played by the constraints in the Eigen theory—the condition of constant organization or, in another version, the condition of constant fluxes.

The Eigen mathematical model is formulated as follows. We have n different types of polymers, their concentrations being denoted by x_i ($i = 1, 2, \ldots, n$). Each type displays metabolism, being capable of self-reproduction and mutagenesis. The kinetic equations

have the form

$$\dot{x}_i = (A_i Q_i - D_i)\, x_i + \sum_{j \neq i} \varphi_{ij} x_j + \frac{\Phi_0 x_i}{\sum_{j=1}^{n} x_j} \tag{17.1}$$

Here $A_i Q_i x_i$ describes the formation of a polymer x_i by way of self-reproduction; $D_i x_i$ describes the degradation of the polymer. Metabolism and self-reproduction are incorporated into the first term. The second term describes mutations—the polymer species i receives extra copies as a result of the corresponding errors in the polymer species j. The presence of mutations is also expressed by the multiplier Q_i which is called the *quality factor*. $Q_i = 1$ implies errorless replication, and $1 - Q_i$ is the measure of errors. Accordingly, the conservation law is valid:

$$\sum_{l=1}^{n} A_l (1 - Q_l)\, x_l = \sum_{l=1}^{n} \sum_{j \neq l} \varphi_{lj} x_j \tag{17.2}$$

Finally, expression (17.1) contains the net flux Φ_0. It is presumed that macromolecules i leave the system or are added to it proportionately, i.e., just as x_i/n, where

$$n = \sum_{j=1}^{n} x_j = \text{const}$$

Moreover, dilution may occur as a result of the influx of the solvent.

Thus, the constancy of the general organization is expressed by the conditions

$$A_i Q_i = \text{const} \quad \text{and} \quad \sum_{j=1}^{n} x_j = n = \text{const} \tag{17.3}$$

The conservation of erroneous copies (17.2) is given by

$$\sum_{l=1}^{n} A_l (1 - Q_l)\, x_l = \sum_{l=1}^{n} \sum_{j \neq l} \varphi_{lj} x_j$$

Accordingly,

$$\sum_{j=1}^{n} (A_j - D_j)\, x_j \equiv \sum_{j=1}^{n} E_j x_j = -\Phi_0 \tag{17.4}$$

since $\sum \dot{x}_j = 0$. The flux Φ_0 compensates for the production of all macromolecules. Equations (17.1) reduce to the equations

$$\dot{x}_i = (W_i - \bar{E}(t))\, x_i + \sum_{j \neq i} \varphi_{ij} x_j \tag{17.5}$$

where

$$W_i = A_i Q_i - D_i \quad (\textit{selective value}) \tag{17.6}$$

$$\overline{E}(t) = \sum_{j=1}^{n} E_j x_j \Big/ \sum_{j=1}^{n} x_j = -\frac{\Phi_0}{n} \quad (\textit{average productivity}) \tag{17.7}$$

Equations (17.5) are nonlinear, since $\overline{E}(t)$ contains all x_i. This term gives a sliding and self-regulating flow of self-organization. The macromolecules that increase in number are those whose selective value W_i is above the threshold $\overline{E}$, i.e., $W_i - \overline{E} > 0$. An increase in their number shifts the threshold $\overline{E}$ to the side of ever increasing values until an optimum value $\overline{E}$ is reached, which is equal to the maximal selective value W_m of all the species available:

$$\overline{E} \to W_m \tag{17.8}$$

Hence, the system will tend to assume a state of "selective equilibrium", which is, however, unstable. It will be disturbed when, as a result of mutations, there will appear a new copy having a higher selective value, i.e., one satisfying the condition $W_{m+1} > W_m$. Following this, the system will pass to a new state of equilibrium, etc. This kind of optimization may be nonmonotonic, since the maximal selective value W_m depends on a given distribution of concentrations x_i, which characterizes the state of the environment.

It is obvious that each mutation leading to a rise in selective value corresponds to a negative fluctuation in entropy production—the increase of order and information. This points to the instability of the existing steady state.

Thus, the kinetic equations having been written, (17.5), we are to solve them. We shall first try to ascertain the significance of the Eigen model from a qualitative standpoint.

The system in question may be visualized as being enclosed in a box with semipermeable walls, through which the monomers and solvent molecules (but not polymers) pass. Inside the box there occur a polymerization reaction, a template replication of macromolecules and their degradation. Macromolecules having the highest selective value are replicated more rapidly and, hence, their number in the box must increase at the expense of the remainder of macromolecules that "die". This is fully supported by computer calculations. Figure 17.1 shows the behaviour of four competing types of macromolecules with different values of W_i. Only those chains "survive" whose selective value is maximal.

Equations (17.5) may be rearranged to

$$\dot{\xi}_i(t) = (\lambda_i - \overline{E}(t))\,\xi_i(t) \tag{17.9}$$

where $\overline{E}(t)$ is given by Eq. (17.7); ξ_l are "normal modes" which are linear combinations of x_i. We have

$$n = \sum_{j=1}^{n} x_j = \sum_{l=1}^{n} \xi_l = \text{const} \tag{17.10}$$

and

$$\overline{E}(t) = \sum_{l=1}^{n} \lambda_l \xi_l(t)/n \tag{17.11}$$

The solution of the system (17.9) at $t \to \infty$ has the following properties. All ξ_l for which $\lambda_l < \overline{E}(t)$ tend to zero, shifting continuously the threshold for selection $\overline{E}(t)$ in accordance with Eq. (17.11) until only one "quasi-species" (usually, a wild type plus a distribution of mutants) is left, which is characterized by a maximal eigenvalue λ_{max}. The selection is described by the extremum principle

$$\overline{E}(t) \to \lambda_{max}$$

The "quasi-species" differs from a "species" with concentration x_m by having not only a dominant form (the "main copy") but all its mutants as well—the normal mode ξ_m corresponding to λ_{max} is a linear combination of concentrations x_m and concentrations of all mutants x_m. The contribution of x_m to ξ_m is predominant.

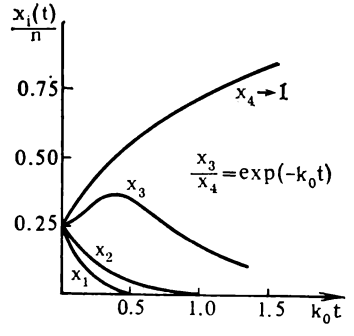


Fig. 17.1. Selection in a system of four competing species:
$W_1 = 1$; $W_2 = 4$; $W_3 = 9$; $W_4 = 10$.

Consider the value of mutations—the quality factor Q_i. The average productivity (17.7) may be represented in the form

$$\overline{E} = \frac{1}{n}\left(E_m x_m + \sum_{l \neq m} E_l x_l\right) = \overline{E}_{l \neq m} + \frac{x_m}{n}(E_m - \overline{E}_{l \neq m}) \tag{17.12}$$

where

$$\overline{E}_{l \neq m} = \sum_{l \neq m} E_l x_l \Big/ \sum_{l \neq m} x_l = \sum_{l \neq m} E_l x_l/(n - x_m)$$

The condition of selective equilibrium $\overline{E} = W_m$ gives the "equilibrium fraction" of the species selected, i.e., its relative survival:

$$\frac{x_m}{n} = \frac{W_m - \overline{E}_{l \neq m}}{E_m - \overline{E}_{l \neq m}} \tag{17.13}$$

The steady-state fraction of erroneous (mutated) chains $1 - x_m/n$ is proportional to $1 - Q_m$:

$$1-\frac{x_m}{n}=\frac{E_m-W_m}{E_m-\bar{E}_{l\neq m}}=\frac{A_m-D_m-A_mQ_m+D_m}{E_m-\bar{E}_{l\neq m}}\equiv\frac{A_m(1-Q_m)}{E_m-\bar{E}_{l\neq m}} \tag{17.14}$$

At $Q_m = 1$, $x_m = n$, i.e., there would have taken place a complete selection of "main copies", but further evolution would have stopped. The evolution requires mutations, i.e., the value of Q_m must be less than unity but larger than a certain value $Q_{\min}$. The value of $Q_{\min}$ is found from the condition $x_m = 0$, i.e., in accordance with Eq. (17.13), $W_m = \bar{E}_{l\neq m}$. We have

$$1>Q_m>Q_{\min}=\frac{D_m+\bar{E}_{l\neq m}}{A_m}=\varepsilon \tag{17.15}$$

The quality factor Q is determined by the accuracy of recognition of the monomer upon template replication. If the recognition of a given unit is independent of the recognition of the other units, i.e., if there is no cooperativity of recognition, the probability of the formation of a correct copy containing ν units is then given by

$$Q = q^{\nu} \tag{17.16}$$

where q is the probability of one unit being exactly reproduced. Comparing Eq. (17.16) with Eq. (17.15), we find that

$$\nu_{\max}<\frac{\ln\varepsilon}{\ln q}\approx\frac{\ln\varepsilon}{1-q} \tag{17.17}$$

since at q close to unity, $\ln q \approx 1 - q$. The inequality (17.17) is the threshold condition which gives the upper limit for the amount of information contained in a stable self-reproducing system.

17.3. Game Models and Informational Aspects of Self-Organization

The key problem being solved by the Eigen model theory is the problem of the emergence of an ordered structure from the original chaos, the problem of creation of information. As we have seen, the appearance of order and its selection and maintenance are possible in an open autocatalytic (template) system that is far from equilibrium.

The accidental self-organization of chaos and the development of irreversible evolution are difficult to visualize. Indeed, the number of various polynucleotide chains having a length of $\nu = 100$ units built up of four nucleotides is equal to $4^{100} \approx 10^{60}$. The accidental

choice of a chain with a certain primary structure has a tiniest probability and, hence, is not realizable.

Without taking into account template self-reproduction, even the eminent physicists (Wigner) have come to the conclusion that prebiotic self-organization is impossible and contradicts the basic laws of physics. We see that this is a seeming contradiction. The Eigen model theory admits self-organization. We shall illustrate the emergence of order from chaos by resorting to game models.

Let us consider the model proposed by Paul and Tatiana Ehrenfest (1907). Suppose we have a chessboard on which black and white draughtsmen are arranged arbitrarily, occupying all the squares. We have two octahedral dice, on one of which are marked the numbers of horizontal rows 1, ..., 8, and on the other the letters indicating the verticals a, ..., h. The rules of the game are as follows: having thrown both dice, we replace a draughtsman, whose coordinates have been obtained on casting, by one of the other colour. The outcome of the game is independent of the initial distribution: nearly equal numbers of black and white pieces will obtain on the board after a sufficient number of throws. This number, which corresponds to one generation, is 64. The Ehrenfest game models the attainment of equilibrium. The probability distribution is Gaussian, with a maximum at $N/2 = 64/2$. In this game a limitation of fluctuations is clearly manifested. If there appears a deviation from the equilibrium distribution, the probability that this deviation will decrease increases in proportion to the deviation. The system is self-regulating in the sense that it returns to the stable equilibrium. This is the general law of the behaviour of the system near equilibrium.

Now we change the rules of the game: a draughtsman chosen by throwing two dice is not replaced by a draughtsman of the other colour; we take another draughtsman of the same colour. In this case, the uniform distribution is unstable—if to the initial state there corresponded 32 white and 32 black pieces, then after 64 throws draughtsmen of the same colour will be left on the board. A chance deviation from the uniform distribution increases and decides the fate of the system. In this game, order arises from disorder as a result of a chance excess in population rather than of the selective advantage for one of the two colours. The game models the "survival of the surviving" and not Darwinian evolution (see page 568).

The third game simulates the behaviour intermediate between the table and the unstable behaviour. This is a game of random walk. This time we cast a coin instead of dice. If a tail obtains, we replace an arbitrary black piece by a white one; if we obtain a head, the replacement is reversed. The uncertainty of an elementary event may have a direct bearing on the macroscopic distribution. The system will oscillate randomly between the two extreme cases (all the pieces are either black or white).

The difference between the three games is demonstrated by the probability values for the extreme cases indicated. The occupation of the board exclusively by black or white pieces in random walk requires, on an average, 32 generations ($32^2 \approx 1000$ throws) if the initial distribution was uniform. In the second game, such a state is attained only after one generation (64 throws). In the Ehrenfest game, for an extreme state to be attained, about 10^{19} throws are needed, since the probability of such a state is $(1/2)^{-64} \approx 10^{-19}$.

In these games there are no conditions for selection: metabolism, self-reproduction (autocatalysis) and mutations. Eigen proposes the "bead game" to illustrate the theory advanced by him.

We have four types of glass beads, say, blue, yellow, red, and green, arranged on the squares of a chessboard. We consider "birth" and "death" to be independent processes. The "death" is simulated by the Ehrenfest game—a bead is removed from the square, whose coordinates were obtained by throwing a die. The result of the next throw corresponds to the second game, i.e., it simulates the "birth"—the bead of the square obtained is doubled, a similar bead is placed on the square that has been freed on the previous throw. The "death" and "birth" alternate, and the total number of beads remains unaltered (64). We now introduce selective advantages into the game. If throwing two octahedral dice gives a "birth", we cast once more an ordinary cubic die and double the blue bead if any number of points obtains on this die; we double the red bead if 1, 2, 3, 4, and 5 points (but not 6) are obtained; we double the yellow bead if 1, 2, 3, and 4 points are obtained, and we double the green if we obtain 1, 2, and 3 points. Thus, the probability of "multiplication" for blue beads is the highest and that for the green beads the lowest. It is essential that in the "birth" phase the dice are cast until the doubling of beads becomes possible. This models regulation by maintaining the total number of beads constant.

The "bead game" is the model of Darwinian evolution. Obviously, only blue beads will eventually "survive"—the entire board will be occupied by them. This is demonstrated by computer calculations. For mutations to be taken into account, we consider a modified version of the game. We have only two types of beads—blue and yellow beads. The blue beads are doubled 4 times faster than the yellow ones. We ascribe to blue beads a certain probability of erroneous reproduction—the replacement of a blue bead by a yellow one. In contrast, the yellow beads multiply without errors—the probability of appearance of a blue bead from a yellow one is negligibly low. Calculations for the error probabilities of 0, 25, 50, and 75 percent show that in the last two cases the blue beads die out; with an error of 25 percent their number exceeds the number of yellow beads, but the yellow beads do not die out; with precise reproduction the blue beads replace the yellow ones. Thus, there is a threshold value for

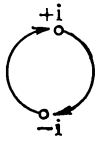
the error percentage at which the predominant multiplication of blue beads is still possible.

The emergence of order in these games implies the creation of information—the memorizing of an accidental choice at the macroscopic level (see Sec. 9.1). In the Eigen model described in Sec. 17.2 we deal with the isolation of the previously available information from the noise that masked it and with the creation of new information. The first process is modelled by the presence of an *a priori* maximal selective value in the "main copies", the second by the appearance of new "main copies" as a result of mutations. Thus, the optimum $\overline{E} \rightarrow W_m$ means the elucidation of information, and the shift $W_m \rightarrow W_{m+1}$ means the creation of new information.

The revealed and created information is left in the system—the system "memorizes" this information. The memorizing of information is a nonequilibrium and irreversible process. We see that the developing system is the creator and receptor of information—the "memorizing" is accomplished by the receptor (see Sec. 17.9). "Memorizing" is a property of a dissipative system. A system close to equilibrium does not "memorize" accidental deviations from equilibrium, fluctuations. On the contrary, a dissipative system is capable of "memorizing" in the sense that it is able to memorize a random choice—to create and receive information.

17.4. Hypercycles

We return to the Eigen model. To make this model be more closely related to nature, we shall assume that macromolecules do not self-reproduce but synthesize complementary chains. The system is described by the alternation of "plus" and "minus" copies ($+i$ and $-i$). Selection occurs among $\pm\, i$ assemblies represented by the cycles



The theory is not changed.

Natural selection among such cycles cannot, however, lead to the appearance of selected long macromolecules. First, the probability of the correct incorporation of a monomer into the polynucleotide chain q (see page 572) does not exceed 0.99. Hence, even at $\nu = 100$ we have $Q = q^{\nu} = (0.99)^{100} = 0.36$. Second, the selection cost expressed by the quantity W_t does not correlate uniquely with the primary structure of the chain. In the chain there may arise mutations and the template synthesis may involve errors without the parameter

W_i being changed. The template replication is based on the recognition of individual units and their nearest neighbours. For example, the double mutational replacement

$$\ldots \text{CABBD} \ldots \text{CACBD} \ldots \rightarrow \ldots \text{CACBD} \ldots \text{CABBD} \ldots$$

must not affect the value of W_i, i.e., the rate of synthesis of the chain as a whole.

A double mutation is, of course, hardly probable. A unit replacement of the link changes W_i, but with a large number of links ν this change is very slight. The relative change of the rate of template synthesis upon replacement of one link is of the order of the quantity ν^{-1}. Suppose that, as a result of the replacement, the selective value W_i has increased by a corresponding small amount. Selection takes some time. If during the time of the predominant survival of the main copies there will appear a second mutation which brings W_i back to the previous value, then no new main copy will evolve. Hence, the quantity W_i does not represent the true *information value* in the case of a simple or complementary reproduction.

No matter how low are the probabilities of mutations, within a sufficient period of time in the system there will accumulate chains that are degenerate with respect to W_m but have different primary structures. If the chains are long, the number of such degenerate macromolecules will be large. In the long run, to the most probable state of the system there will correspond the maximum number of different primary structures of chains with equal values of W_m. Hence, the selective equilibrium is not only unstable but it must inevitably degenerate.

For macromolecules to undergo evolution, an autocatalytic system is required that makes use of the information stored in the chain, i.e., a system reproduced by recognizing not individual links but the primary structure of the chain as a whole or a considerable part of it.

What has been discussed above refers to polynucleotide chains. We see that any apparently long chains (according to the estimate by Eigen, with $\nu > 100$) do not undergo evolution by themselves.

Protein macromolecules contain a larger amount of information than the corresponding polynucleotides, since they possess different tertiary (and quaternary) structures. But proteins are not capable of autocatalysis, self-reproduction. They contain a large amount of information that has a low selective value.

Eigen arrives at the conclusion as to the evolutionary necessity of the autocatalytic hypercycle built up of elementary polynucleotide cycles $\pm i$ (page 575) with the catalytic participation of proteins synthesized with the aid of the same polynucleotides.

Consider a sequence of reactions in which the products of each of the steps undergo further transformation. If in such a sequence some

product coincides with the reactant of the preceding step, the system forms a reaction cycle, which is a catalyst as a whole. As a matter of fact, we have already encountered a catalytic cycle in considering enzymatic reactions (Fig. 17.2). Another example is the citric acid cycle or the Krebs cycle (page 442).

The Eigen catalytic hypercycle consists of information carriers—the polynucleotide cycles I_1, I_2, ..., I_n, which instruct their own reproduction and translation during the synthesis of enzymes E_j, which perform the role of polymerases—they serve as catalysts for the reproduction of the next information carrier I_{j+1} (Fig. 17.3).

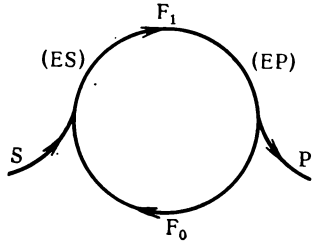


Fig. 17.2. The cycle in an enzymatic process.

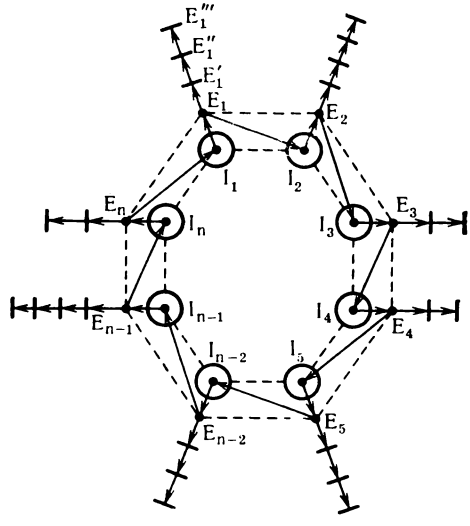


Fig. 17.3. A self-instructing catalytic hypercycle:
I_j —information carriers, i.e., $\pm j$-cycles of RNA; E_j—enzymes coded by j and catalyzing the generation of carriers I_{j+1}; E'_j and E''_j—enzymes catalyzing parasitic processes.

The hypercycles make use of the entire information stored in polynucleotide chains or a considerable part of it. The hypercycle has the following properties:

1. It provides a stable and controlled coexistence of all species united through cyclic linkages.
2. The hypercycle provides a coherent growth of all its participants.
3. The hypercycle competes with any individual replicative unit that does not belong to it.
4. The hypercycle is capable of increasing or decreasing in size if this change gives a selective advantage.
5. The hypercycle may unite into networks of a higher order.
6. The hypercycles compete with each other. The inner linkages and cooperative properties of the hypercycle may develop up to an optimum function. "Phenotypic" advantages, i.e., variations, which

give a distinct advantage to a mutant, are immediately stabilized. On the other hand, "genotypic" advantages, which favour the emergence of the next product, require a spatial separation for a competing fixation.

For hypercycles the selective advantages are always functions of the size of population due to nonlinearity. Therefore, once established, the hypercycle is not easy to replace by a new one, since a new species always appears in a small number of copies. The information fixed in the hypercycle cannot be easily lost.

These properties of hypercycles are illustrated by model numerical calculations. Hypercycles that contain a sufficient number of links are characterized by limit cycles on the phase portraits. All the components of the hypercycle coexist, experiencing correlated nonlinear oscillations. This provides a mutual stabilization of the components.

The Eigen theory has been compared with the experiments on the "evolution in a test tube" carried out by Spiegelman (1965-1971) and Weissmann (1976). In these experiments the mutants of the $Q\beta$-phage were studied. The Eigen theory models well their selection.

How could hypercycles, systems that are already sufficiently complex, have emerged? As a matter of fact, this is the basic question pertaining to the Eigen theory. According to Eigen and Schuster, the first self-reproducing structure with a stable content of information is a molecule of the type of tRNA built up of about 100 nucleotides. This supposition had been made earlier by Kuhn (1972) (see Sec. 17.5). These molecules of the pre-tRNA were selected as a quasi-species. They were stable towards hydrolysis, since they had not only a secondary but also a tertiary structure (page 289). This structure was inherently chiral. A few such molecules having similar functions but different specificities were required for constructing a translation system. Such a system may have arisen from a single quasi-species, but equivalent partners must have been simultaneously developed. This cannot be done by uniting molecules into a self-reproducing macromolecule of a larger size because of the error threshold. Nor is this possible via compartmentation because of a severe competition inside a compartment. Hence, the process requires that functional relationships be established among all the units. These relationships must admit a competition with erroneous copies, otherwise the information could not have been preserved. These relationships must "switch off" the competition between the units that are to integrate into a new functional system and provide their cooperation. And, finally, the integrated functional system must successfully compete with any other less efficient systems. These three requirements can be fulfilled in hypercycles alone.

In the early stages of evolution, which are characterized by inaccurate replication and translation and by a low content of efficiently

replicating units, the hypercyclic organization offers substantial advantages.

Eigen and Schuster have proposed concrete models for the origin of the genetic code. In particular, arguments are proposed to support the view that the start of translation is favoured by primary structures enriched in hypercycles. Translation arose due to a sufficient amount of proteinoids, protein-like substances (see page 567). The resulting translation system could have been improved, but it did not admit a new translation system with an alternative configuration. In this way, a unified genetic code was created and improved. The subsequent stages of evolution involved compartmentation and genetic recombination, the quality factor Q and the maximum number of links in the functional chains, ν_{max} being increased during the course of evolution.

These considerations are reasonable, but the problem of the origin of a hypercycle, a genetic code, and a translation apparatus remains open in the Eigen theory. We cannot presume that these events occurred by chance, with a very low probability. Had this been the case, the very origin of life itself would have had a vanishingly low probability. We look for a model interpretation of prebiotic and biological evolution that would account for their legitimate origin.

17.5. Other Models of Prebiological Evolution

Let us first consider the thermodynamic aspects of evolution. The Eigen model, which accounts for the selection of the pre-tRNA with a highest selective value, describes the mechanism of the fluctuational decrease of entropy in a dissipative system. A mutation corresponding to a rise in selective value leads to an increased order. This is shown schematically in Fig. 17.4.

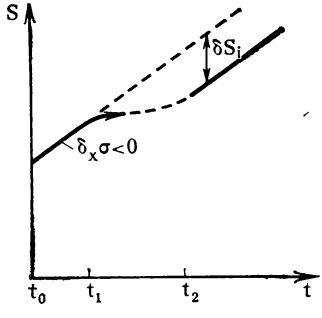


Fig. 17.4. The time dependence of entropy S for the process of selection at constant total fluxes of monomeric units. The time interval t_0t_1 represents a stationary state (σ = constant); the interval t_1t_2 gives the region of instability; at $t > t_2$ a new steady state appears.

Prigogine and his coworkers (1972) have conducted a phenomenological treatment of evolution in terms of the theory of stability.

Suppose we have n interacting substances X_i ($i = 1, 2, \ldots, n$) with sufficiently high concentrations x_i. For an open system we have

$$\dot{x}_i = \mathcal{F}_i^e(x_1, \ldots, x_n) + \mathcal{F}_i(x_1, \ldots, x_n) \qquad (17.18)$$

The first term describes the flux of the substance from the surrounding medium and the second, the reactions taking place inside the system. We assume $\mathcal{F}_i^e$ to be constant and the system to be homogeneous. Suppose that Eqs. (17.18) have at least one asymptotic stable solution; this means that all n roots of the characteristic equation for the system have negative real parts.

Let us now assume that there accidentally arise mutants of substances X_i, i.e., their erroneous copies Y_j ($j = 1, 2, \ldots, s$) with concentrations y_j. The order of Eqs. (17.18) increases by s. If $s = 1$ (one mutant), the characteristic equation of the increased system will then contain an additional term of order ε. At small ε the characteristic equation will have $n + 1$ roots, where n roots are close to the roots of the system (17.18). The stability of the steady state of the increased system will be influenced only by a new root λ_{n+1}, which may depend on ε either as

$$\lambda_{n+1} \sim \varepsilon^{-1} \tag{17.19}$$

if the correction term is a multiplier in the term containing an $(n + 1)$st power of λ, or as

$$\lambda_{n+1} \sim \varepsilon \tag{17.20}$$

if the appearance of Y_j adds a constant term to the characteristic equation.

As a result, in both cases there may appear roots with a positive real part. A small perturbation of the system (17.18) may lead to a substantial change in its stability. In the case of (17.19), this change will be rapid, and with (17.20) it will be slow. In the case of (17.19) the evolution of Y is described by the equation

$$\varepsilon \dot{y} = G(\{x\}, y, \varepsilon) \tag{17.21}$$

where $\{x\}$ is a set of all the variables x_i. Equations (17.18) take the following form

$$\dot{x}_i = \mathcal{F}_i^e(\{x\}) + \mathcal{F}_i(\{x\}, y, \varepsilon) \tag{17.22}$$

In the case of (17.20), instead of Eqs. (17.21) and (17.22), we get

$$\dot{y} = G_1(\{x\}) + \varepsilon G_2(\{x\}, y, \varepsilon) \tag{17.23}$$

$$\dot{x}_i = \mathcal{F}_i^e(\{x\}) + \mathcal{F}_{1i}(\{x\}) + \varepsilon \mathcal{F}_{2i}(\{x\}, y, \varepsilon) \tag{17.24}$$

Let us consider the case (17.19). We introduce a new variable $\tau = t/\varepsilon \gg t$ and rewrite Eq. (17.21) in the form

$$\frac{dy}{d\tau} = G(\{x\}, y, \varepsilon) \tag{17.25}$$

Equations (17.21) and (17.22) describe the evolution of the system in various time scales. Equations (17.21) and (17.25) represent the fast production of a new substance Y at practically constant values of $\{x\}$, and Eqs. (17.22) represent the gradual change of $\{x\}$ after y has reached its steady-state value, corresponding to $G(\{x\}, y, \varepsilon) = 0$. Thereby, selection must occur rapidly. The situation is thermodynamically equivalent to the one presented in Fig. 17.4—the fluctuational entropy production is negative.

In the case of (17.20), Eqs. (17.23) and (17.24) describe (at an instability) the gradual evolution of substance Y to a new regime. Comparison with Eq. (17.6) shows that the factor ε may be treated as the average rate of mutations, the instability being directly determined by copying errors. In the case of (17.19), the role of mutations consisted in introducing relationships between the old and new variables. The evolution through instability was determined by the emergence of new elements, say, new enzymes.

In Eq. (17.23) we put $G_1 = 0$, i.e., we assume that the new substance is produced by mutations alone. The original system may become unstable if G_2 is an increasing function near $(\{x_0\}, y, 0)$, where x_0 represents the solutions of the initial system.

These conceptions are nicely illustrated by kinetic models that describe autocatalysis and phase transitions—the change from the production of one species of polymers to that of another species of polymers.

In 1972 Kuhn advanced a spectacular model of prebiotic evolution which takes direct account of the circadian periodicity in the state of the environment. Such a periodicity implies a periodic change of temperature and humidity—the alternation of the dissolution of the substance and the drying of the solution. There were first formed relatively short polynucleotide chains, the pre-tRNA's, which acquired a tertiary structure. The periodic change in the states of the environment acted as a selection factor—only those molecules were preserved which had not been hydrolyzed in the humid stage. Chains were selected that were capable of replication. This was accompanied by the appearance of chirality, whose sign was determined by an initial chance event (see page 48).

These processes took place at the boundary between the sea and the land, molecules were synthesized in the pores (compartments) of clay minerals, into which energized monomers found their way. The periodicity of the state of the environment determined the alternation of the multiplication of molecules and their selection, the alternation of convergent and divergent phases. This was followed by extension, i.e., the chains became more and more complex, and at each stage the forms capable of multiplication were functionally tied up with the environmental conditions. A stepwise expansion of life space began to evolve.

The next stage, according to Kuhn, consisted of the formation of associates from molecules of the pre-tRNA. Only those molecules were selected that were structurally related to one another. The associates were practically correctly multiplied, despite the frequent errors in some molecules. What occurred is, in fact, similar to the operation carried out during the industrial assembly of cars from separate parts. The parts are controlled in advance, separately, prior to assembly.

Kuhn emphasizes the importance of self-assembly not only in contemporary living systems but also in prebiotic evolution.

During the subsequent evolution, according to Kuhn, the system became more and more independent of the rather specific environment as a result of its increasing complexity. The informational aspects of the increase of complexity will be discussed in Sec. 17.10. This was also accomplished by way of alternation of divergent and convergent phases. The convergent phase implies the detailing of the existent organization and the divergent phase, the rearrangement of the system and the creation of new information. As a result of such rearrangement, the life space was expanded. The associates increased in size, penetrating into larger-pore regions. Later, they served as catalysts (templates) for the synthesis of a second species of macromolecules (proteins), which built compartments, blocking the pores in minerals and hindering the diffusive separation of template molecules. There was developed a feedback between the polynucleotides responsible for the synthesis of polypeptides and these polypeptides. Membranes were formed which made the system independent. There evolved a genetic code as a by-product of the evolutionary development. Accordingly, there appeared primitive enzymes.

The subsequent stages involved the rearrangement of the genetic apparatus—the separation of the replication mechanism and the mechanism of enzyme synthesis. This led to the appearance of a third, latest species of macromolecules—polydeoxynucleic acids (DNA).

Having reached a certain level of complexity, such systems encounter a limit of the genetically transferrable amount of information. It is about 10^7 bits. The optimum value of error upon replication of a nucleic acid is 10^{-6}. This is the system-bound limit. It was surmounted in nature through the emergence of sex and genetic recombination. The evolution of a sex apparatus already required exceedingly refined structural elements—a kind of explosive evolution of cell organelles. This was the switch from prokaryotes to eukaryotes. The switch was jumpwise (similar to a phase transition!), since no intermediate systems existed.

The Kuhn theory offers simple estimates of the time required for the successive phases of prebiotic and biological evolution (see

Sec. 17.6). As we have seen, the hypercycle in this model is formed later than membranes.

Chernavskaya and Chernavsky remark that the supposition of the presence of specific enzymatic (polymerase) functions in randomly synthesized polypeptides in the Eigen theory and in the Kuhn theory is unsatisfactory. This supposition introduces into the model an event whose probability is exceedingly low. Chernavskaya and Chernavsky propose a model of prebiotic evolution different from those described. This model admits the possibility of protein synthesis on a polynucleotide molecule just as on a heterogeneous catalyst. This synthetic pathway itself predetermines the shape of the protein molecule and, hence, its function.

The assembly on the surface of a double-helical polynucleotide gives rise to a complementary structure of the protein that forms a coat for the polynucleotide. The jacket has a replicase function if the complementary correspondence between the coat and DNA (RNA) is not complete and a mechanically strained conformation arises in the complex. An important role here may be played by a periodic change of the environmental conditions. The resulting nucleic acid-protein complexes are no longer capable of self-reproduction, which, however, fails to provide the fixation of "positive" features: a sheath may be synthesized on the daughter double helix which hinders replication. It is possible that primary adaptors of the type of tRNA arise in the system. With incomplete complementarity the system contains grooves into which there penetrate molecules that are complementary to the double helices, on the one hand, and to one or a few amino acids, on the other. The adaptor has the conformational flexibility required. Such a mechanism may serve as the primary mechanism of translation which appears before the evolution of the genetic code.

The model is given in mathematical form by means of appropriate (nonlinear) kinetic equations and of an examination of the phase patterns.

The emergence of the genetic code may be determined by two mechanisms. First, in accordance with the Eigen theory, there is possible the preferential survival of objects with a protein coat, which provides the highest rate of replication. Such a selection may give rise to an assembly of objects with identical protein coats but with different nucleotide sequences. To one amino-acid sequence there will correspond several nucleotide sequences. In this case, however, evolution may be interrupted as a result of "reverse degeneration". There is possible the formation of adaptors complementary to identical nucleotide sequences and to different amino-acid blocks. This leads to errors in protein synthesis. According to Chernavskaya and Chernavsky, the reverse degeneration is overcome by antagonistic interactions. When molecules with different nucleotide sequences

interact with one another, their sets of adaptors may be mixed. This will interfere with protein synthesis in objects that have encountered, and they both will perish; this will not take place in encounters of identical objects. The kinetic (dimensionless) equations in the simplest case of two objects, x and y, have the form

$$\frac{dx}{d\tau} = x - xy, \quad \frac{dy}{d\tau} = y - xy \tag{17.26}$$

The phase portrait of the system is shown in Fig. 17.5. The isoclines (see page 506) of the horisontals are straight lines $y = 0$, $x = 1$, and the isoclines of the verticals are straight lines $x = 0$, $y = 1$. The system has two unstable steady states—the unstable node $x = 0$, $y = 0$ and the saddle $x = 1$, $y = 1$. The system proves to be operating by the all-or-none law: depending on the initial conditions, the trajectories run either to the point $x \to \infty$, $y \to 0$ or to the point $x \to 0$, $y \to \infty$. Thus, one of the two possibilities is chosen.

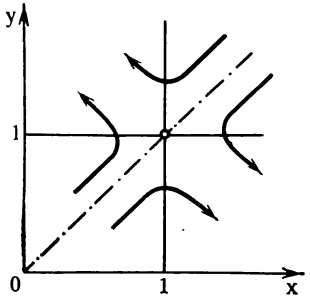


Fig. 17.5. The phase pattern of the system (17.26).

The occurrence of antagonistic interactions during prebiotic evolution nevertheless requires more rigorous argumentation.

Ebeling and Feistel (1979) have proposed a model which combines the Eigen theory with the compartmentation concepts—the coacervates according to Oparin. The compartments arose in the early stages of chemical evolution. They gave rise to chemical *microreactors*. The driving force of evolution was the competition between these microreactors, in which there evolved a "molecular language" that later led to the evolution of the genetic code. Catalytic cycles and networks were built in such microreactors. Introducing space inhomogeneity, Ebeling and Feistel overcome the difficulties associated with the formation of the hypercycle. Coacervate-microreactors play the role of the first primitive cells capable of division and containing the precursors of the evolution of more complex molecular and supermolecular structures.

The models described above cannot be regarded as reproducing the actual processes of prebiotic evolution. They show, however, that the physico-mathematical modelling of prebiotic and biological evolution is basically possible. They allow one to come to a conclusion of fundamental importance for theoretical biophysics:

contemporary physics is, in principle, sufficient for the understanding of biological phenomena, among which the most important are the phenomena of development.

The investigations in this field are in the initial stage, the models described are informative and instructive, though of transient nature. However, the physical substantiation of theoretical biology has been started.

17.6. Biological Evolution

The acquaintance with the living world and with the foundations of the theory of evolution immediately poses numerous questions before the investigator. If evolution is not directional at all and consists, at each stage, of the trial-and-error sorting of all accidentally arising mutations with the predominant survival of mutants best adapted to the environmental conditions, then how could such complex living systems have evolved for a time period of 3.5×10^9 years? How could, in general, complex and perfect organs, such as, for example, the eye of vertebrates with its regulatory devices or the facet eye of an insect, have evolved at all by way of natural selection? It might seem that for natural selection there was neither sufficient time nor sufficient material—a diversity of mutants.

Such are the fundamental questions that face the physicist speculating on evolution. Darwin, who constructed the theory of evolution, yet failed to provide answers to these questions. The emergence of a highly ordered biosphere from chaos remained obscure on the basis of pure statistics. It is for this reason that there appeared finalistic theories of directed evolution—from the Berg nomogenesis to the Omega principle of Teilhard de Chardin.

In the last several decades the situation has changed owing to a deeper general understanding of the evolution of the Universe and to the successes achieved in genetics and molecular biology. Though no strict quantitative theory of evolution is available, we can give positive answers to the question indicated: yes, both the time and the material were sufficient.

As Mayr states (1968), the modern concept of the world is based on the knowledge that the Universe, stars, the Earth, and all living creatures have evolved during the course of a long history which had not been predetermined or programmed, the history of continuous gradual changes corresponding to more or less directional processes compatible with the laws of physics. As has already been said above, cosmic and biological evolution are similar in this respect (page 16).

Darwin's theory is based on four main principles:

1. The world is not static but is in the process of development.
2. The evolution is continuous and gradual.

3. The diversity of species, including the biosphere as a whole, have a single common ancestor—the life on the Earth originated in a quite definite way.

4. The mechanism of evolution is natural selection, a two-stage process consisting of the appearance of variant types and of the survival of the fittest in the course of the struggle for existence.

According to Darwin, a given species is represented by a principal genotype with a small number of rare variants which deviate from it. The discovery and the first studies of mutations have, it might seem, confirmed this concept, since the probabilities of mutations are low.

In actual fact, as we know at present, each population contains an enormous reserve of genetic variants, which are not at all manifested directly in the differences between macroscopic features. It has been established that in populations of any significant size there are genetic variants for nearly all organismic features. Gel electrophoresis has made it possible to discover a wide variability of proteins, sets of *isoenzymes*, which characterize genetic variants at the molecular level. This is also evidenced by a practically unlimited diversity of *lymphocytes* and *antibodies* in man and higher animals (see Sec. 17.8). The variability of proteins expressed by the fraction of heterozygous loci per species is 17 percent in plants, 13.4 percent in invertebrates, and 6.6 percent in man and vertebrates. The last figure is the lower limit.

The extreme genetic diversity is determined by a number of factors, the most important of which are *genetic recombination* due to sexual reproduction and the presence of predominantly *neutral mutations*, which do not affect life activity. The majority of genetic variants in populations arise not from new mutants in each generation but as a result of the "shuffling" of the previously accumulated mutations by way of recombination.

The genetic variability is actively maintained by a number of factors, primarily by the heterozygous advantages—individuals that are heterozygous with respect to many loci are usually stronger and can be more successfully reproduced than homozygous individuals.

Of course, the factors responsible for gene mutations operate randomly in the sense that these mutations arise independently of their future adaptability to the environment. Finalism in evolution is completely rejected by the fact that the original changes in the genetic material may be only mutational, i.e., random.

Thus, there was more than enough material for evolution—at least from the time when sexual reproduction appeared. But was there sufficient time?

The random, chance appearance of mutations does not imply the absence of directionality in evolution. Such directionality arises in a natural way, since selection through the survival of the fittest is

ı directional rather than an accidental factor. We have already seen hat this factor may have operated in prebiotic evolution (see Sections 17.2 through 17.5). The creation of new information is the nemorizing of a random choice. The selection is random but the nemorizing process is not random, since it is determined by the onstraints imposed by the environment.

Let us note in passing that sexual reproduction—genetic recombinations—implies a continual generation of new information, since he combination of genotypes of two given heterosexual individuals s accidental and not programmed in advance.

The directional evolutionary progress therefore arises automatically, because evolution favours any variant type that gives advantages n competition. The successful operation of natural selection does ccur and is possible owing to the high individualization of biological ystems. The evolution cannot repeat itself precisely because these ystems are individual and any problem posed by the environmental onditions has numerous solutions. The presence of this multitude f solutions implies that the state of a system undergoing evolution s unstable. Mutations cause a transition of the system to one of the elatively stable states. The instability determines the directional, rreversible character of evolution.

Special emphasis should be placed on the nonuniformity of the volution of various structures of the organism. Regulatory and tructural genes undergo evolution at different rates. Enzyme systems, etc., also undergo evolution at different rates.

How did evolution progress in time? Geology and paleontology provide a wealth of information on this problem.

The age of the Earth is 4.5×10^9 years. Chemical evolution began bout 4×10^9 years ago and continued for more than 2.5×10^9 ears—up to the formation of the present-day oxygen atmosphere. hemical and biological evolution overlapped—the first prokaryotes ppeared 3.5×10^9 years ago. The entire Precambrian period, which took 2.5×10^9 years and ended 1×10^9 years ago, was needed for he emergence of the first eukaryotes—single-celled algae. Multicellular organisms appeared for the first time only 0.5×10^9 years go, land plants and vertebrates evolved 0.4×10^9 years ago, mamnals about 0.2×10^9 years ago, the primates 0.05×10^9 years ago, nd *Homo erectus* about 0.02×10^9 years ago. Thus, the greater art of the time was expended on the development of the first primitive cells, following which the evolution progressed at a faster ate.

How did such complex systems evolve and why were these systems ot destroyed by mutations?

One of the key propositions of modern evolutionary theory is that nce created the genetic information is preserved and does not dispear. It is exactly this factor that is responsible for the formation

of certain species of plants and animals. This factor is also responsible for *recapitulation*—the similarity between the embryos of different groups of vertebrates.

The formation of complex organs is accounted for not by finalistic but by the directional character of evolution. It is dictated by the regularities of ontogenetic development which embodies a certain genotype realized in accord with environmental conditions. In this sense, *convergent phenomena* are characteristic, such as, for example, the formation of the eye similar to vertebrate eye of the cephalopod mollusks which represent a different branch of evolution.

The physico-mathematical treatment of evolution encounters a number of difficulties. Darwin's followers have no doubt in the fact that evolution is consonant with the laws of physics but it is pointless to state that evolution is reduced to physical laws. Biological evolution is the result of specific processes that take place in specific systems, the explanation for which becomes reasonable only at the level of complexity of these processes and systems. Nonetheless, it is fundamentally possible to construct physico-mathematical models of evolutionary processes. Mathematical population genetics is a well-developed branch of science.

The main difficulty in the understanding of evolution is associated with the erroneous conception that the probabilities of sequential evolutionary events—the accumulation of mutations and variations of features—are independent of each other. In fact, we are dealing here with *Markov complex chains*—the probability of a given event depends on the realization of this or that event in the entire sequence of the preceding stages. A Markov simple chain is characterized by the so-called *stochastic matrix*, the members of which p_{ij} express the probabilities that an event i occurred at the nth step if at the $(n-1)$st step there occurred an event j. If the chain is complex, the stochastic matrix itself is continuously changed. The results of evolution exert an effect on the boundary conditions—the environmental conditions. A population that undergoes evolution is itself a part of biogeocenosis, in which the evolution of the given population and of the populations of other species occurs. The directionality of evolution is expressed by the fact that many members of the changing matrix are equal to zero.

An important role in divergent evolution (adaptive radiation) is played by the mechanism based on the association with subsequent mutations. Examples of association are symbiosis, diploidization, the combination of DNA portions. It is only due to association that from a single small precursor with a limited length of DNA there could have evolved diverse species with a considerably longer DNA chain.

Identical organisms, associates, are capable of giving rise, as a result of mutations, to different organisms with different proteins. The interaction of such different organisms will no longer be strongly

antagonistic (cf. page 584), since the sets of adaptors in them are the same. Association is essential at all levels of the biological structure, at all stages of evolution.

The physico-mathematical description of evolution must be based on the achievements of population genetics, on the general theory of nonlinear dynamical and stochastic systems. Each new stage of evolution is the result of the instability of the preceding state of the system. As has already been said (page 532), the emergence of a new species is similar to a phase transition.

We shall return to biological evolution in Sec. 17.9.

17.7. Ontogenesis

In contrast to phylogenesis, the directionality of the development of an individual—*ontogenesis*—is immediately obvious. Ontogenesis includes three basic processes—*cell differentiation*, *growth of cells*, i.e., the increase of the number of cells and of the mass of the embryo, and *morphogenesis*, i.e., the development of certain organs and whole organisms. All these processes occur in accordance with the genetic program.

Cell differentiation is based on molecular differentiation. All the somatic cells of a multicellular organism contain the same set of genes which is identical to the genome of the original zygote. This proposition has been proved by direct experiments. Gurdon has established that if the nucleus of an epithelial cell of a tadpole's intestines is introduced into an egg of a frog, whose nucleus has been preliminarily destroyed by ultraviolet radiation, a normal individual is developed from such an egg. In a specialized cell most of the genes are repressed; only certain proteins can be synthesized in it. When a specialized genome is transferred to a zygote, the genes become free from repression and normal development takes place. Chemical studies of cell differentiation have led to important results. Wessels and Rutter grew a culture of an embryonic tissue—the pancreatic gland of a rat. This organ produces enzymes that participate in food digestion and hormones that regulate carbohydrate metabolism. The so-called exocrine cells secrete zymogenes—the precursors of digestive enzymes, and the endocrine cells secrete hormones—glucagon and insulin. They studied the production of all these proteins at the various stages of embryonic development and arrived at a correlation between the protein content and the appearance of certain intracellular structures. The production of specialized proteins increases with time not monotonically but passes through three regulatory phases (Fig. 17.6). It has been found that the mesodermal cells from other organs promote the differentiation of the pancreatic cells. This is presumably accounted for by the action of a certain factor, perhaps of a protein nature.

The synthesis of secretory proteins in the culture of the epithelial cells of the pancreas is specifically suppressed by actinomycin D, an antibiotic that blocks the synthesis of RNA. The sequential suppression of the synthesis of various proteins by actinomycin shows that the corresponding mRNA's are synthesized at different times (cf. page 308).

Perhaps, the developmental phases are determined by the cooperative action of numerous genes. Some groups of genes are activated and other groups are simultaneously inactivated; no successive activation of each of hundreds of individual genes takes place. It is

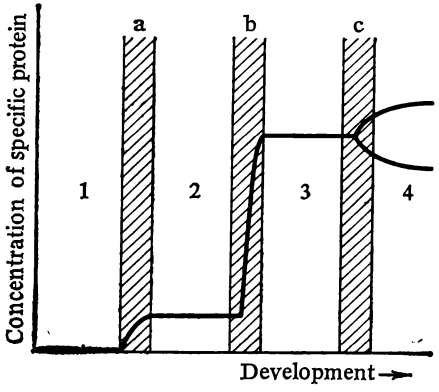


Fig. 17.6. The change of the products of specialized proteins in the course of embryogenesis:
1—predifferentiated state; *2*—protodifferentiated state (step I); *3*—differentiated state (step II); *4*—definitive state (step III); *a*, *b* and c are transitions.

possible that the concerted cooperative transitions, which display the features of phase transitions (see below), are governed by changes in the chromosomal structure.

Ontogenesis involves two types of processes. The first type includes regulatory intra- and intercellular interactions brought about by the phenomena of molecular signalling, molecular recognition. It is these phenomena that largely determine differentiation. The second type includes active movements of cells as a result of mechanochemical processes, which are also stimulated by molecular signalling. These movements determine *morphogenesis.*

The cell differentiation itself, beginning with the appearance of the first two blastomers, is the result of intracellular interactions, regulation of the activity of genes by the substances of the cytoplasm and cell membrane. Differentiation at the early stage (the *blastula*) is determined by two factors of the very general character. The first of these is the nonuniform distribution of the substance of the cytoplasm of the original zygote, and the second is the inhomogeneity of the medium inside the cellular sphere produced as a result of cell division. Either of them implies the presence of *positional infor-*

mation (Wolpert). Apart from these factors, ontogenesis is also determined by contact and humoral regulation.

In the later stages of development, hormones begin to operate. A complicated mode of development—metamorphosis in insects—is controlled by *ecdysone*, a steroid hormone that regulates the growth of the organisms.

One of the most important substances in ontogenesis is cyclic AMP (see page 44), which serves for molecular intra- and intercellular signalling. It has been shown that cAMP is capable of inducing differentiation and morphogenesis in undifferentiated cells *in vitro*.

The role of cAMP in development is spectacularly demonstrated in the biology of *myxomycetes*. If there is more than enough food, these organisms exist as amoeba-like unicellular organisms. With a food deficiency myxoamoebas aggregate into a pseudoplasmodium, which develops as a multicellular organism containing tens of thousands of cells, until conversion to sporophores and spore multiplication occur. Aggregation and differentiation of myxomycetes occur as a result of the liberation of cyclic AMP and reception of these molecular signals.

Ontogenesis is similar to phylogenesis—the evolutionary development—in the sense that a developing embryo is a sort of biocenosis (ecosystem), in which there coexist populations of specialized cells. The "embryonic biocenosis" itself is formed by these populations. The "embryonic biocenosis" is a dynamical system with a predetermined behaviour and not a static one. Dividing cells are not liable to random variations. The development of biocenosis is governed by external influences and by intra- and intercellular interactions, but it is programmed by a genotype. We shall discuss the informational aspects of ontogenesis in Sec. 17.9.

The tasks of theoretical modelling of ontogenesis are still far from being solved. However, a number of promising works in this field deserve attention.

It is natural that the first attempts at a model description of ontogenesis were purely biological. Gurvich has proposed a formal method of treatment of intercellular interactions involved in morphogenesis, imparting to each cell some "*biological field*" which spreads beyond the cell and affects its neighbours. The fields of several cells are vectorially added and cause a movement of cells required for shape formation (1944). Waddington has introduced the concept of an *epigenetic landscape*. The development of an organism is likened to the movement across a rugged terrain in phase space. The route of movement is determined by the character of the terrain and by external factors. In ontogenesis there occurs a channelling of development—the morphogenetic potentials of cell populations are gradually limited. Waddington coined the term *creode* for a channelled development trajectory which attracts the nearest ones (1968).

The physico-mathematical modelling of ontogenesis began with

the work of Turing mentioned earlier (page 524). In essence, this work solves a problem which is more complicated than required: the emergence of a space-time order in a quite homogeneous dissipative system. Indeed, in the original zygote there may occur the polarization of the material; the zygote is not homogeneous at all (amphibians). This is determined by the process of generation of an egg cell (ovum) in an inhomogeneous organism. A cycle is realized: an inhomogeneous zygote is produced from an inhomogeneous organism, and in turn an organism arises from a zygote. The theoretical modelling of ontogenesis on this basis has not yet been carried out.

The basic propositions of present-day model theories based on the works of Turing are as follows:

1. The position of the cell in a developing system is characterized by its radius-vector r_i.

2. The concentration gradients for functional substances are sufficiently small for the size of the cells to be neglected. Accordingly, r_i may be regarded as being continuously changed.

3. The state of the cell is defined by a set of concentrations of functional substances, $x_j(r_i)$.

4. The equations for the material balance describe chemical kinetics plus diffusion (see page 524):

$$\frac{\partial x_j}{\partial t} = \mathcal{F}_j(x_1, x_2, \ldots, x_n) + D_j \frac{\partial^2 x_j}{\partial r_i^2} \tag{17.27}$$

where n is, say, the number of various proteins, being of the order of 10^2.

5. Different macroscopic structures correspond to different stable steady-state solutions.

The geometric, topological problems may be considered separately from kinetics. Wolpert (1969) points out that in a biological system it is the path of development that is programmed but not the structure of the emerging organism. The path of formation of a geometrical structure, which is axial in the simplest case, i.e., linear, may be formalized. The problem consists in explaining the regulatory preservation and restoration of the structure when it has been partially destroyed. For example, hydroids, which have an axial organization, restore their structure after removal of some part of it. Wolpert has proposed a topological "three-coloured flag model", which offers a phenomenological qualitative explanation of the regulation of development based on the consideration of a linear series of interacting elements, each of which may be in several alternative states.

Martinez (1972) has considered a model consisting of a linear array of cells. A cell in the array is capable of division only if the concentration of some substance in the cell reaches a threshold level. The rate of change of this concentration depends on the difference in the concentrations of the other two substances, "morphogenes",

one of which may diffuse between the cells. The Martinez model, which is a modification of the Turing model, leads to a nonuniform distribution of matter among the cells and to their specific division.

The theory of systems that have multiple steady states has been worked out by Lavenda (1972) in application to biosynthetic processes.

We shall proceed from the operon model (see Sec. 8.8). The corresponding scheme is presented in Fig. 17.7. The regulator gene (RG)

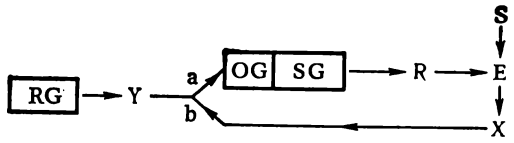


Fig. 17.7. Schematic representation of biosynthetic processes. See the explanation in the text.

responsible for the synthesis of repressor gene Y, which shuts down the operation of the operator gene (OG) and, hence, of the structural gene (SG). The structural gene produces mRNA R, which codes the synthesis of enzyme E, which transforms substrate S into metabolite X capable of being reversibly bound with Y. The repressor Y is an allosteric protein, which may have various conformations having different activities in binding with the operator gene. The branches *a* and *b* in Fig. 17.7 represent the active and inactive forms of Y. The transitions between them depend on the concentration of X. The rate of synthesis of mRNA is limited by the binding of X by the repressor Y. If Y has n active sites for X, the degree of saturation of the repressor with metabolite, i.e., the average number of X molecules bound by the Y molecule, is equal to

$$\bar{X} = \frac{\sum_{i=0}^{n} iL_iX^i}{\sum_{i=0}^{n} L_iX^i} = \frac{\partial \ln \left(\sum_{i=0}^{n} L_iX^i \right)}{\partial \ln X} \qquad (17.28)$$

where $L_i = \prod_{j=0}^{i} K_j$, where K_j is the equilibrium constant of the reaction

$$YX_{j-1} + X \rightleftharpoons YX_j$$

$L_0 = K_0 = 1$ (cf. page 234).

The kinetic equation for the synthesis of mRNA has the form

$$\dot{R} = k_1 X - k_2 R \tag{17.29}$$

Moreover,

$$\dot{E} = k_3 R - k_4 E \tag{17.30}$$

$$\dot{X} = k_5 S E - k_6 X \tag{17.31}$$

For the sake of simplicity, it is assumed that the loss of R, E, and X is determined by the monomolecular degradation. Equation (17.29) depends on the nonlinear relationship between Y and X, i.e., on the cooperative properties of Y. The simplest scheme that gives a dynamic cooperative behaviour is obtained for a repressor with three binding sites ($n = 3$). In this case, the steady-state values of X satisfy the equation

$$X_0^3 - (3\gamma - L_2/L_3)\, X_0^2 + (L_1 - 2L_2\gamma)\, L_3^{-1} X_0 - (\gamma L_1 - 1)\, L_3^{-1} = 0 \tag{17.32}$$

where

$$\gamma = \frac{k_1 k_3 k_5}{k_2 k_6} S$$

The dependence of the rate of synthesis and breakdown of mRNA on the concentration of X is given in Fig. 17.8. The necessary con-

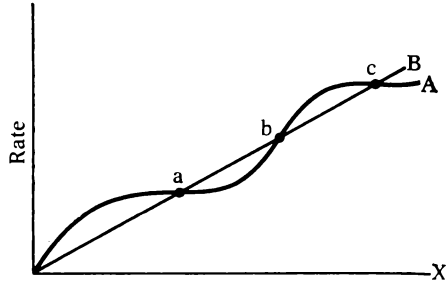


Fig. 17.8. The dependence of the rate of synthesis (A) and decomposition (B) on X: a, b, c—steady states.

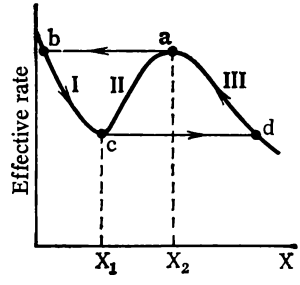


Fig. 17.9. The dependence of the effective rate of synthesis on X: a and c—bifurcation points; b and d—stable states.

dition for the existence of three steady states is $3K_3 > K_1 > 2K_2$. The difference between the constants K_1, K_2, and K_3 and, hence, between L_1, L_2, and L_3, implies cooperativity. Dividing Eq. (17.28) by X, we obtain the effective rate of synthesis and the rate of breakdown of mRNA (Fig. 17.9). The values of X_1 and X_2 are determined by the condition

$$X_{1,2} = (\gamma - L_2/3L_3) \mp [(\gamma - L_2/3L_3)^2 - (L_1 - 2L_2\gamma)/3L_3]^{1/2} \tag{17.33}$$

These are bifurcation points. The branches *I* and *III* are stable, and the branch *II* is unstable. At the bifurcation points there are attained the threshold values of concentration, which, along with the critical concentration gradients, determine the regulation of the system.

The transitions between the various steady states are similar to phase transitions (Sec. 15.5). The multiplicity of steady states can also be determined directly by the behaviour of enzymes having different conformational states and, hence, jumps in catalytic activity.

The regulatory, triggering properties of a system having multiple steady states are demonstrated by a model of two criss-crossed operons (Jacob and Monod, 1964; Fig. 17.10). The metabolite X_1 produced

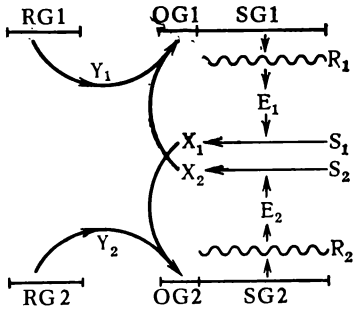


Fig. 17.10. The Jacob-Monod trigger circuit.

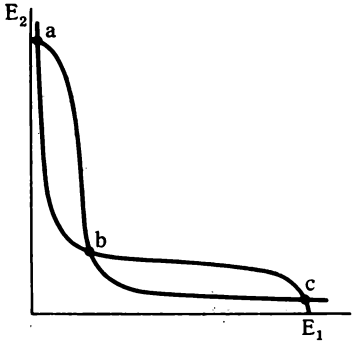


Fig. 17.11. The steady-state curves for a trigger system.

by the first operon is bound by the repressor Y_2 of the second operon and vice versa. The kinetic equations in the absence of cooperativity in repressors Y_1 and Y_2 have the form (A, B, and k_i are constants)

$$\dot{R}_1 = \frac{A}{B+E_2} - k_2 R_1, \qquad \dot{R}_2 = \frac{A}{B+E_1} - k_2 R_2$$
$$\dot{E}_1 = k_3 R_1 - k_4 E_1, \qquad \dot{E}_2 = k_3 R_2 - k_4 E_2 \tag{17.34}$$

Under steady-state conditions $\dot{R}_1 = \dot{R}_2 = \dot{E}_1 = \dot{E}_2 = 0$ and

$$E_1 = \frac{Ak_3}{k_2 k_4} \frac{1}{B+E_2}, \qquad E_2 = \frac{Ak_3}{k_2 k_4} \frac{1}{B+E_1} \tag{17.35}$$

These two curves intersect only at one point and the system is not a trigger system. However, if instead of Eqs. (17.34), we consider a

nonlinear, cooperative system of equations, which in the simplest case has the form

$$\dot{R}_1 = \frac{A}{B+E_2^2} - k_2 R_1, \quad \dot{R}_2 = \frac{A}{B+E_1^2} - k_2 R_2$$
$$\dot{E}_1 = k_3 R_1 - k_4 E_1, \quad \dot{E}_2 = k_3 R_2 - k_4 E_2 \tag{17.36}$$

trigger properties will appear. The steady-state curves are expressed by the formulas

$$E_1 = \frac{Ak_3}{k_2 k_4} \frac{1}{B+E_2^2}, \quad E_2 = \frac{Ak_3}{k_2 k_4} \frac{1}{B+E_1^2} \tag{17.37}$$

These curves are shown in Fig. 17.11. There are three points of intersection, a, b, and c, of which a and c are stable and b is unstable. Such a system can switch, with a slight change in the parameters, from state c, in which the protein E_1 is synthesized, to state a, in which the protein E_2 is synthesized.

The modelling of ontogenesis is carried out exactly on the basis of the multiplicity of the steady states and trigger properties of systems that have dynamical instabilities. The consideration of such systems based on analogies with phase transitions is important and promising.

17.8. Immunity

Specific processes of cell differentiation are responsible for immunity, in particular, for the production of antibodies in the organisms of vertebrates.

The structure and functions of antibodies, their interaction with antigens have been described in Sec. 4.8. Let us now consider the phenomenon of immunity.

The immune system serves to protect animals from the harmful effects of microorganisms. However, this is not the whole story: immunity provides control of the genetic constancy of the cells of an organism. The main function of the immune system is to remove mutant (in particular, cancer) cells from the organism of an animal. In response to the introduction of an antigen the organism produces specific reactive cells (cell-mediated immune response) and specific antibodies (humoral immune response). Both reactive cells and antibodies circulate in the organism and interact specifically with antigens. As a result, the foreign material is inactivated, destroyed or phagocytized by the cells of the reticuloendothelial system. The present-day conceptions of immunity are based on the *Burnet clonal-selection theory* (1949). The organism produces lymphocytes, each of which is sensitive to a single antigen or to several related antigens. This gives rise to lymphocytes which are sensitive to

nearly all antigens, including those the organism has never encountered in biological conditions. This is determined by the presence on the membranes of lymphocytes of specific receptors, which exhibit high affinity for a certain antigen. An antigen acts as a selection factor which stimulates the development of clones of immunologically active cells from lymphocytes that have receptors to the given antigen.

Antibodies are generated by mature plasmatic cells (*z*-cells). The antibody specificity coincides with the specificity of receptors that reside on the surface of cell precursors. Such are so-called B-cells which belong to the category of small lymphocytes and are produced as a result of the differentiation of stem blood cells. The role of receptors in B-cells is played by immunoglobins (10^4-10^5 molecules per cell). B-cells become capable of division and proliferation after being transformed to so-called *blast cells* (*y*-cells) under the influence of antigens. Blast formation occurs with the lapse of the latent period which lasts for 24-48 hours. The *y*-cells undergo intensive proliferation. Part of *y*-cells give rise to clones of plasmatic *z*-cells. A clone of *z*-cells produces antibodies. Mature *z*-cells do not divide, their lifetime being several tens of hours.

Apart from giving rise to a clone of *z*-cells, antigen-stimulated B-lymphocytes are capable of leading to the formation of immune-memory cells. The secondary immune response of the organism to the antigen that has earlier acted on it usually proves to be more rapid and stronger than the primary response. This phenomenon is called *immune memory*; it is associated with an increase in the number of cells capable of responding to a repeated antigenic stimulus, just as is the case with the original B-lymphocytes.

The second population of lymphocytes capable of specific recognition of genetically foreign materials is the so-called T-cells. Like B-cells, they are produced from stem cells, but during their development they pass through the thymus. Some T-cells are responsible for cell immune responses. In response to an antigenic stimulus they differentiate into specific reactive killer cells. The other two subpopulations of T-cells—helper cells and suppressor cells—play a regulatory role in the development of humoral immunity. There also exist thymus-independent antigens capable of inducing a normal response without the help of T-cells.

Figure 17.12 is a schematic representation of the events described.

In immunity phenomena we encounter a number of fundamental problems of biology and biophysics. These are the problems of cell differentiation and molecular and cellular recognition. The immunity phenomena have been studied relatively well.

We shall dwell here on the model theory of humoral immunity developed in the works of Dibrov, Livshits, and Volkenshtein (1976-1978). This theory takes into account the delay of the immune

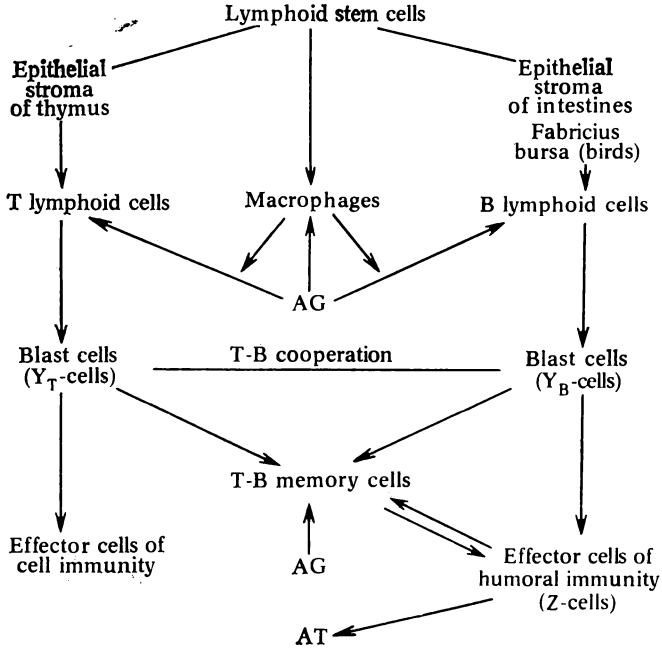


Fig. 17.12. Schematic representation of events caused by the injection of an antigen.

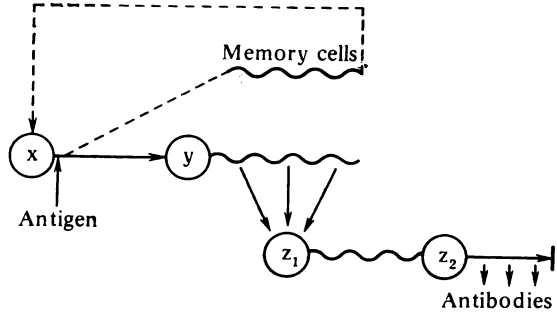


Fig. 17.13. Diagram depicting the proliferation of AT-producing cells. The wavy lines denote proliferation.

response, i.e., the production of antibodies, with respect to the moment of stimulation. The development of z-cells from B-lymphocytes takes 3-4 days. For memory cells to be formed a longer time period is needed. The scheme of development of a clone of antibody-producing cells is shown in Fig. 17.13.

A mathematical analysis shows that the delay time may be decisive for the character of the immune response.

The rate of antibody production is proportional to the number of plasmatic cells present, which is determined by the number of B-lymphocytes stimulated earlier by antigens, at the moment of time separated from the current time by the delay time t_r. The production of memory cells is analogously determined by the number of acts of stimulation of B-lymphocytes at the moment of time separated from the current time by the delay time t_m.

The process is described by a system of differential equations with a delaying argument:

$$\begin{aligned}
\dot{x} &= J - \tau^{-1}x(t) - Px(t)\,g(t) + A_m x(t-t_m)\,g(t-t_m)\,\theta(t-t_m)\\
\dot{g} &= Kg(t) - Qh(t)\,g(t)\\
\dot{h} &= A_r x(t-t_r)\,g(t-t_r)\,\theta(t-t_r) - Rh(t)\,g(t) - Sh(t)
\end{aligned} \tag{17.38}$$

Here x, g, and h are, respectively, the numbers of B-lymphocytes, antigens (say, multiplying microorganisms), and antibodies;

$$\theta(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geqslant 0 \end{cases}$$

The quantity J is the rate of increase of population of B-lymphocytes as a result of the differentiation of stem cells; τ is the average lifetime of B-lymphocytes. The second and third terms on the right-hand side of the equation for $\dot{x}$ describe, respectively, the decrease of the number of B-cells as a result of the contact with antigen and its increase due to the formation of memory cells, which are considered to be identical with the original B-lymphocyte; K is the rate constant for antigen reproduction in the organism; the terms with Q and R describe, respectively, the decrease of the number of antigens and antibodies as a result of their interaction. The term incorporating A_r describes antigen production; S represents the rate of degradation of antigens. The parameters P, A_m, A_r, t_m, and t_r may depend on the dynamics of populations of regulatory T-cells. This dependence may be ignored in dealing with thymus-independent antibodies as well as in the case of the rapid attainment of a steady-state number of T-cells. These parameters are assumed to be constant.

For a detailed study of immune responses it is expedient to simplify the system (17.38). Antibodies display high affinity for antigens. Therefore, as far as the number of antigens is not too small, the antibodies produced react rapidly with antigens. Indeed, appreciable amounts of antibodies are observed only upon elimination of antigens. This allows us to approximate the system (17.38) by the following equations:

$$\begin{aligned} \dot{x} &= J - \tau^{-1} x(t) - P x(t)\, g(t) + A_m (t - t_m)\, g(t - t_m)\, \theta(t - t_m) \\ \dot{g} &= K g(t) - A_r \frac{Q}{R}\, x(t - t_r)\, g(t - t_r)\, \theta(t - t_r) \end{aligned} \tag{17.39}$$

These equations are similar to the Volterra equations for the prey-predator system (see Sec. 15.3) if $A_m > P$. The difference is that here account is taken of the delay in the nonlinear terms.

The inspection of Eqs. (17.39) shows that the character of steady-state points on the phase patterns depends significantly on the delay. For rapidly multiplying antigens to be efficiently suppressed the time t_m must be neither too small nor too large. One might expect that the increase of t_m must always lead to an increase in the time needed for the elimination of antigens. In actual fact, however, the increase of instability with increasing delay may lead to a reduction of the reaction time because of a decrease in the number of rounds about the steady-state point.

We shall now take into account the final rate of antibody-antigen interaction, considering that the number of B-lymphocytes changes only slightly in the course of the immune response, i.e., $x(t) \approx \approx x(0)$. Equations (17.38) now become

$$\begin{aligned} \dot{g} &= K g(t) - Q h(t)\, g(t) \\ \dot{h} &= A g(t - t_r)\, \theta(t - t_r) - R g(t)\, h(t) - S h(t) \end{aligned} \tag{17.40}$$

where $A = A_r x(0)$. If $KR/AQ < 1$, the system has two steady-state points (0, 0) and $h_0 = K/Q$, $g_0 = (S/R)(KR/AQ)/(1 - KR/AQ)$. The point (0, 0) is unstable, and the character of the second point depends on the parameters and, in particular, on the delay time t_r. Figure 17.14 shows the phase trajectories of the system (17.40) in terms of the dimensionless variables g/g_0, and h/h_0 for $KR/AQ = 0.5$, $S/R = 3$ and different values of the delay $t_r K$. The steady-state point in Fig. 17.14*a*—a stable focus—corresponds to such a course of the disease when, beginning from a certain moment of time, the disease stops progressing, and the amount of antigen in the organism is constant—the organism becomes a carrier of infection. The low value of g_0 corresponds to recovery. Figure 17.14 (*b*, *c*, *d*) shows limit cycles corresponding to the periodic progress of the disease. As the lag increases the amplitude of oscillations

increases and the minimal amount of antigen is reduced. The time fraction during which the amount of antigen is close to zero increases. The case depicted in Fig. 17.14*e* corresponds to instability and an unlimited increase of antigens, i.e., to the death of the organism.

If the final rate of the antigen-antibody interaction is taken into account within the framework of the deterministic equations (17.38) through (17.40), the complete elimination of multiplying antigens

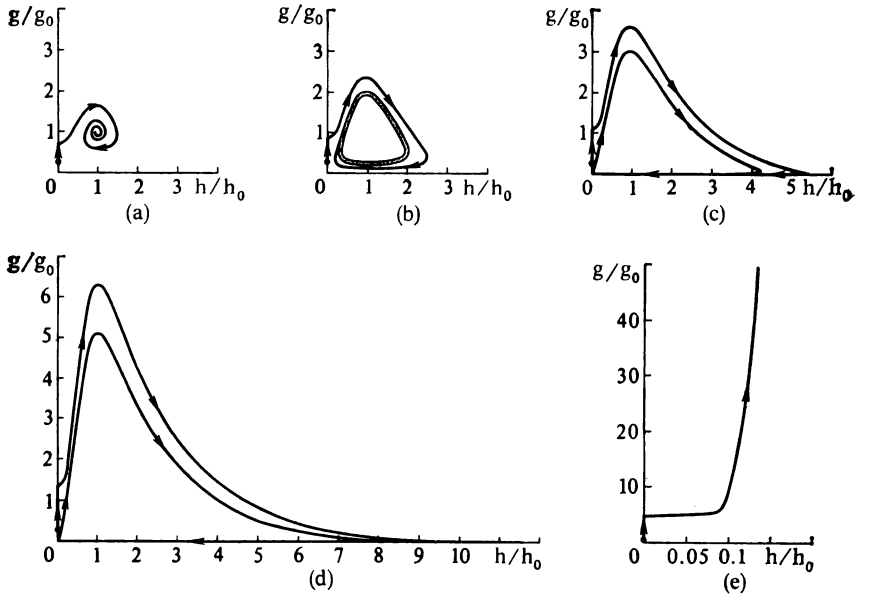


Fig. 17.14. The phase trajectories of the system (17.40) at lag values of $t_r K$ equal to 1 (*a*), 1.25 (*b*), 1.75 (*c*) and 3 (*d*).

is impossible. But at small antigen concentrations one has to consider a discrete stochastic process (see Sec. 15.6) and antigens can be completely destroyed.

The theory allows one to substantiate the tactics of curing an infectious disease. It is expedient to inject an immune serum (antiserum), i.e., specific antigens at the moment of the maximum amount of innate antigens. The strongest effect in the use of antibiotics is possible if they are injected at the moment when the amount of antigens is at a maximum. If antibiotics and antigens are introduced too early, the innate immunity response of the organism is reduced, which increases the probability of a relapse. In the cases of weak chronic

infections one of the curing procedures is to deliberately bring the disease up to an acute form, which makes it possible to activate the immune response of the organism and, thereby, to increase the probability of the complete elimination of antigens.

The problems of the origin of immunity are developmental problems. We may expect that the principles underlying the treatment of immunity, such as taking account of the delay and also the consideration of phase transitions in nonequilibrium systems with multiple steady states, will contribute to the modelling of ontogenesis, carcinogenesis, and biological evolution.

17.9. Biological Evolution and Information Theory

As pointed out earlier, biological development is associated with the decrease of the entropy of a developing system and, hence, with the increase of information (see pages 575 and 587). The increase of the amount of information occurs either through isolating the information from the noise that masks it or creating new information as a result of the memorizing of an accidental choice (Quastler). For such a choice to be possible, the system must have multiple steady states.

The phase space of the system contains stable and unstable regions. The latter are crossed by separatrices (see page 510). To the stable regions there corresponds information and to the unstable region the entropy. The conversion transformation of part of entropy to information in a thermodynamic process, say, during the crystallization of a liquid, implies the entry of the system into the region of phase space which is the only stable one under given conditions. Here new information does not appear—the formation of a crystal is uniquely programmed by the structure of its molecules.

In a thermodynamic process the information increases but it is not new but *a priori* known information hidden earlier by the "entropy noise". A system that creates information, i.e., a system with multiple steady states, and picks out one state from several states, is not a thermodynamic system; it is a dissipative system. The appearance of new information during evolution is connected, in particular, with the "memorizing" of randomly chosen neutral mutations and genetic recombinations, with the formation of associates (page 582). New information is created in each crossing (interbreeding) process, during the emergence of a new individual. Systems that create information are the Universe as a whole, the Earth, the biosphere, and the man in his creative activity. Speciation, which occurs according to Darwin by way of divergent evolution, implies the generation of new information.

Schmalhausen has carried out an important work on the translation of the theory of evolution into the language of information theory.

Biocenosis receives information about the state of populations by means of certain feedback mechanisms and, hence, "incorporates a specific mechanism of transformation of this information to control signals and means of transmission of these signals to the population". The corresponding scheme is shown in Fig. 17.15. The inherited information is passed on from generation to generation only after its

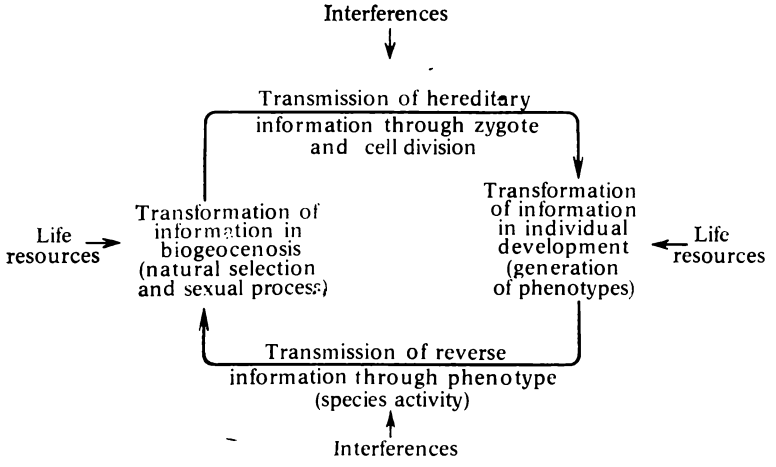

Interferences
Transmission of hereditary information through zygote and cell division
Life resources
Transformation of information in biogeocenosis (natural selection and sexual process)
Transformation of information in individual development (generation of phenotypes)
Life resources
Transmission of reverse information through phenotype (species activity)
Interferences

Fig. 17.15. Schematic representation of the regulating mechanism of evolution according to Schmalhausen.

transformation in the biogeocenosis, a part of which is the population. The information about changes in the population is transferred by way of multiplication of species, whose progeny enters into the same population.

Schmalhausen has, however, pointed out that not the amount of information contained in the chromosome, cell, species, population, is essential for evolution, but the *quality* or *value* of information. "In all cases when information is compared and selected, this is accomplished on the basis of its quality evaluation... In biology the qualitative evaluation of information is of primary importance".

We shall make an attempt to characterize the value (quality, sense) of information.

The value of information in a message is determined only by the results of the receipt of this message by a certain system. The universal determination of the value of information (similar to the determination of its amount; page 314) is impossible—what is valuable for one receptor system may be of no value for another. Thus, the

value of information is directly associated with its reception. The receipt of information is a nonequilibrium and irreversible process of transition of the receptor system from an unstable state to one of possible stable states. The receipt of information thus implies the creation of new information in the receptor system. Such systems differ in reception levels. At a given level only nonredundant information is of value. *Redundant information* is repetitive information, the transmission of which lowers the probability of suppressing information by noise. Suppose we have a text written in Russian. If we know only that the Russian alphabet contains 32 letters, the amount of information per letter will be equal to

$$I_1 = \log_2 32 = 5 \text{ bits}$$

At the next level of reception we know the probabilities of appearance of letters, and the amount of information per letter is expressed by the Shannon formula (page 315)

$$I_2 = -\sum_{i=1}^{32} p_i \log_2 p_i$$

The probabilities p_i are normalized to unity. We have $I_2 < I_1$ and at the second level of reception the message has a redundancy equal to

$$R_2 = 1 - I_2/I_1$$

At the subsequent reception levels we take into account double, triple, etc., letter correlations. The redundancy

$$R_j = 1 - I_j/I_1 \tag{17.41}$$

increases with the reception level. According to Shannon, for the written English language we have

I_1	I_2	I_3	I_4	...	I_6	...	I_9
4.76	4.03	3.32	3.10	...	≈ 2.1	...	≈ 1.9 bits

and the redundancy

R_1	R_2	R_3	R_4	...	R_6	...	R_9
0	0.15	0.30	0.35	...	≈ 0.56	...	≈ 0.60

At the ninth level 60 percent of letters contained in the message are redundant, the remaining 40 percent being sufficient for the receipt of the message. The increase of redundancy implies the increase of the indispensability of nonredundant letters.

The statement made by Schrödinger that the organism is an aperiodic crystal is not only the assertion of the condensed state and aperiodicity. An aperiodic system contains a much greater amount of valuable nonredundant information than an equivalent periodic system.

Thus, the information value is determined by the nonequilibrium nature and instability of the receptor system and by the levels of irreversible reception, only the nonredundant information being valuable. We may define the quantitative value of information as the increase of the probability of attainment of a certain objective as a result of the message being obtained.

The value of information may be defined tentatively as the "degree of nonredundancy", i.e., the irreplaceability of the message. Suppose that a message contains N_1 "letters" and the amount of information is $N_1 I_1$. At the next level the information received is reduced to $N_1 I_2$ ($I_2 < I_1$). We replace the decrease of information per letter by the decrease in the number of letters in the message to $N_2 < N_1$. This is equivalent to the neglect of the redundant letters. The gross information is assumed to be constant. Then

$$\frac{N_2}{N_1} = \frac{I_2}{I_1} = 1 - R_2$$

The information value is assumed to be accordingly increased. Then the ratios of the values at the successive steps of reception are equal to

$$\frac{1}{1-R_1} : \frac{1}{1-R_2} : \ldots : \frac{1}{1-R_n} \quad \text{or} \quad 1 : \frac{I_1}{I_2} : \frac{I_1}{I_3} : \ldots : \frac{I_1}{I_n}$$

In the case of the English language (page 604) these relative values are equal to

$$1 : 1.18 : 1.43 : 1.54 : \ldots : 2.28 : \ldots : 2.50$$

Consider the reception of genetic information coded by the nucleotide sequence in DNA. At the level of reception of this text the amount of information in the DNA chain containing n nucleotides is equal to

$$I_1 = \log_2 4^n = 2n \text{ bits}$$

This information is valuable only for the translation process which takes place in the nonequilibrium system of the cell. The amount of information in a synthesized protein chain is given by

$$I_2 = \log_2 20^{n/3} = 1.44n \text{ bits}$$

and the redundancy caused by the degeneracy of the code is

$$R_2 = 1 - I_2/I_1 = 0.28$$

At the next level we take into account the replacements of some amino acids by related ones with the properties of the protein remaining unchanged. If the number of indispensable amino acids is $N < 20$, then

$$I_3 = \log_2 N^{n/3} < I_2$$

The redundancy at this level is given by

$$R_3 = 1 - I_3/I_2 > R_2$$

At the fourth level we take into account the possibility of replacement of proteins, say, proteases. If two species of DNA chains containing n_1 and n_2 units are introduced, proteins with $n_1/3$ and $n_2/3$ units are synthesized. Let $n_1 = 0.75n_2$ and $N = 16$ for both proteins. Then

$$I_1 = 2n_1 + 2n_2 = 4.67n_1 \text{ bits}$$
$$I_2 = 1.44n_1 + 1.44n_2 = 3.36n_1 \text{ bits}$$
$$I_3 = 1.33n_1 + 1.33n_2 = 3.10n_1 \text{ bits}$$

and, if the proteins are interchangeable,

$$I_4 = 1.33n_1 \text{ bits}$$

The values of the redundancies R_4, R_3, R_2, and R_1 are equal to 0.72, 0.34, 0.28, and 0. The relative values are as follows: 3.56, 1.50, 1.39, and 1.00.

The value of an element of informational message may be defined as the degree of its irreplaceability.

Let us now determine the values of codons according to the degree of irreplaceability of the amino acids coded by them. Using experimental data, Bachinsky (1976) has established the values of the *functional similarity* of amino-acid residues (AAFS) which characterize their interchangeability. This functional similarity is defined as

$$\text{AAFS} = \frac{2N_{ij}}{N_i + N_j}$$

where N_{ij} are the total number of positions in isofunctional proteins at which some representatives have amino acid i and the others have amino acid j; N_i and N_j express the frequency of occurrence of the residues. Table 17.2 presents the values of N_i, N_{ij} (in the left bottom half) and AAFS (in the right upper half of the table).

The conditional values of codons are determined by considering the results of their point mutations. Let us illustrate this by an example. We have the codon AAA (Lys). The mutations and the corresponding values of AAFS are as follows:

CAA Lys-Gly 18	ACA Lys-Thr 19	AAC Lys-Asn 17
GAA Lys-Glu 17	AGA Lys-Arg 25	AAG Lys-Lys 100
UAA Lys-Term 4	AUA Lys-Ile 5	AAU Lys-Asn 17

To the replacement by an identical residue (silent mutation) we arbitrarily assign the value of AAFS = 100. We sum up the given values and divide the sum by 300. We obtain the tentative value of q, which is the larger, the greater is the interchangeability of the residues. For AAA, $q = 0.74$. The conditional value, which charac-

Table 17.2. Values of N_i, the Matrix N_{ij} and AAFS

		N_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	Ala	455		29	27	17	9	13	16	18	17	10	29	12	6	12	8	4	7	3	6	4	1
2	Ser	430	129		31	10	7	16	25	13	18	7	25	13	12	12	4	7	5	4	5	4	3
3	Thr	345	109	122		16	9	19	17	17	14	12	15	15	13	9	4	6	6	4	7	6	2
4	Val	338	69	40	55		23	10	7	9	5	35	13	12	7	9	8	10	5	6	11	2	0
5	Leu	332	37	25	29	78		6	6	6	2	24	5	10	5	4	16	7	5	1	15	4	1
6	Lys	308	50	58	62	32	19		17	17	14	5	11	18	25	9	1	5	7	0	5	0	4
7	Asn	287	60	32	54	21	20	50		15	27	3	20	15	11	6	4	7	14	1	3	2	1
8	Glu	284	66	48	52	27	20	51	41		31	5	14	25	8	7	1	5	4	2	3	0	1
9	Asp	262	61	63	44	14	7	39	75	86		3	18	12	6	8	0	5	6	1	2	0	0
10	Ile	236	35	24	36	99	67	13	8	14	7		5	8	7	4	10	11	4	1	10	4	1
11	Gly	232	99	84	43	37	15	31	52	35	44	11		12	8	11	6	5	4	2	2	5	2
12	Gln	230	41	43	44	35	27	49	39	63	29	19	27		15	10	2	4	12	2	4	3	0
13	Arg	227	21	40	36	19	14	66	27	20	15	16	18	35		5	5	8	7	1	6	3	2
14	Pro	188	39	38	25	23	11	23	15	17	17	8	23	21	10		3	3	3	1	1	1	3
15	Phe	159	26	13	10	20	40	3	10	3	1	19	12	3	9	5		28	9	1	5	13	2
16	Tyr	141	12	20	15	23	17	12	14	10	10	21	10	7	14	5	42		9	1	5	6	0
17	His	139	20	15	15	11	11	16	29	8	13	7	8	23	13	5	13	13		2	2	6	0
18	Cys	104	9	12	9	14	3	1	2	4	1	2	6	3	5	1	1	1	2		2	0	0
19	Met	99	17	13	16	25	32	11	5	6	4	16	3	7	9	1	7	6	2	2		4	0
20	Trp	48	9	9	4	3	8	0	3	0	0	5	3	4	4	1	13	6	6	0	3		0
21	Term	33	2	7	3	0	1	6	1	3	2	1	2	0	3	3	2	0	0	0	0	0	

terizes the degree of irreplaceability of the residue, is the greater, the smaller is the value of q. We assume the information value to be equal to

$$v = (q + 0.5)^{-1}$$

and for the example given $v = 0.81$. Table 17.3 gives the values of v obtained in this way. These values have the meaning of tentative averaged characteristics expressing the sequence of the values of the codons.

Table 17.3. Values of Codons xyz

x \ y	A		C		G		U		z
A	Lys	0.81	Thr	0.55	Arg	0.70	Ile	0.64	A
	Asn	0.79		0.55	Ser	0.80	Ile	0.65	C
	Lys	0.81		0.57	Arg	0.70	Met	1.25	G
	Asn	0.79		0.55	Ser	0.80	Ile	0.65	U
C	Gln	0.86	Pro	0.60	Arg	0.51	Leu	0.49	A
	His	0.94		0.61		0.61		0.57	C
	Gln	0.86		0.60		0.51		0.50	G
	His	0.94		0.61		0.61		0.57	U
G	Glu	0.76	Ala	0.52	Gly	0.57	Val	0.53	A
	Asp	0.77		0.52		0.55		0.54	C
	Glu	0.76		0.52		0.57		0.55	G
	Asp	0.77		0.52		0.55		0.54	U
U	Term		Ser	0.56	Term		Leu	0.68	A
	Tyr	0.98		0.56	Cys	1.12	Phe	0.86	C
	Term			0.54	Trp	1.82	Leu	0.69	G
	Tyr	0.98		0.56	Cys	1.12	Phe	0.86	U

We see that the values of the jointly degenerate codons may differ appreciably, since the results of their mutations are different. This shows that silent mutations are not unimportant for evolution. Averaging v over all jointly degenerate codons, we find the tentative values of amino-acid residues in proteins which characterize their degree of irreplaceability (Table 17.4).

The most valuable and, hence, irreplaceable is the residue Trp. As we have seen, this is also valid for terminal mutations (page 305).

Biological development, both ontogenesis and phylogenesis, progresses in the direction of increasing complexity of the biological

Table 17.4. Values of Amino-Acid Residues in Proteins

	v		v		v
1. Trp	1.82	8. Lys	0.81	15. Pro	0.60
2. Met	1.25	9. Asn	0.79	16. Leu	0.58
3. Cys	1.12	10. Asp	0.77	17. Gly	0.56
4. Tyr	0.98	11. Glu	0.76	18. Thr	0.55
5. His	0.94	12. Ile	0.65	19. Val	0.54
6. Gln	0.86	13. Ser	0.64	20. Ala	0.52
7. Phe	0.86	14. Arg	0.61		

system. The increase of complexity implies the increase of the number of different elements of the system and of relationships between them. The irreplaceability of the elements of the system is increased and, hence, the information value is also increased (see the next section).

The increase of the information value in biological development is evidenced by a number of facts. Recapitulation means the increase of the irreplaceability of information in embryogenesis. At the early stages of ontogenesis this manifests itself very distinctly. A certain region of the blastula or early gastrula of a newt is responsible for the formation of the eye. This region is the *presumptive eye*. If it is transplanted into the embryo which is in its later stage of development, the fate of the presumptive eye will depend on the transplantation site—it can form a brain or an eye on the head of the animal and, in the other regions, those organs and tissues that are characteristic of these regions in normal development. At the neurula stage the presumptive eye becomes the *determined eye*—the corresponding region forms an eye upon transplantation into any locus of the embryo. *Totipotency* is replaced by *unipotency* in the course of development, i.e., the irreplaceability is increased. Similar processes occur during phylogenesis and speciation.

The irreplaceability of the element of the system implies the decrease of its adaptability when an irreplaceable element is put out of operation, the organism may perish, since the function of this elements cannot be performed by another element. However, the increase of the complexity of the system due to the increase of the irreplaceability of its elements makes the organism more autonomic and, hence, increases its adaptability.

One might think that in biological development the information value must increase, i.e., the irreplaceability of the elements of the message even at the simplest molecular level of the structure must increase. Let us make use of the values of v given in Table 17.4

and calculate the differences of the overall values of cytochromes *c* of a number of organisms (Table 17.5).

Table 17.5. Differences in the Total Values of Cytochromes *c*

Man	0.00	Whale	—0.88	Penguin	0.00
Rhesus monkey	—0.10	Kangaroo	—0.88	Hen	—0.05
Donkey	—0.34	Dog	—1.06	Duck	—0.30
Horse	—0.43	Elephant	—1.22	Emu	—0.30
Pig	—0.58	Bat	—1.25	Dove	—0.58
Rabbit	—0.66			Tortoise	—0.80

On the basis of these data we may infer that the value of the protein chain increases during phylogenesis. Of course, we must not think that the composition of cytochrome *c* can reflect the greater perfection of the human brain as compared with the brain of an animal. The greater value means that the man has travelled a longer path of evolution.

During the course of biological development not only the information value present in the organism increases but also the ability of the biological system to select valuable information. This ability reaches a higher level in higher animals, whose sense organs are designed for such selection. A frog responds only to a moving insect; a bat, using ultrasound location, receives only reflected, but not direct, signals. The selection of valuable information underlies the entire creative activity of man.

As we have seen (page 318), in order to obtain one bit of information, an energy no less than $kT\ln 2$ must be used. This value is the cost of one bit of any information—both valuable information and valueless information, which is not indispensable. The selection of valuable information does not require additional energy expenditure. For instance, the channels in the membrane of a receptor cell, through which molecules or ions of certain shape and size can pass, will suffice. The energy associated with the formation of the specialized channels or receptor sites of membranes has been consumed at the earlier, preceding stages of biological evolution.

Thus, a living system is capable of selecting valuable information, without any special payment. In other words, the living system is developing, approaching a state in which it "pays" $kT\ln 2$ for one bit of valuable information and does "pay" nothing for nonperceivable valueless information.

The conceptions expounded here require further development and fill-in of details. One might expect that the amino-acid residues in a protein globule differ in irreplaceability—value. In the region of an active site the residues are more indispensable. Outside the active site the replaceability is high and the value is small. According-

ly, in this region there occurs a random drift of replacements, rapid evolution based on neutral mutations or mutations that slightly reduce adaptation (the Kimura neutralistic theory; see page 124). The informational concepts will undoubtedly prove useful for the further development of the theory of evolution. The creation of new valuable information is possible only in the presence of a pool of information of low value.

17.10. Complexity and Evolution

Having considered the informational aspects of biological development, we shall now, in conclusion turn to the concept of complexity intimately connected with information theory. It is a commonplace to speak of the increase of the complexity of biological systems during the course of evolution. Whatever the definition of complexity, it is obvious that a multicellular organism is more complex than a unicellular organism, that man is more complex than a blue-green algae. Science, however, requires rigorous definitions.

Ordinarily, while speaking of complexity, we imply the number of components that make up a system. Such a definition appears, for example, in the classic work of von Neumann devoted to the theory of self-reproducing automata. John von Neumann pointed out that with a high complexity of an object its description may be more complex than the object itself. In complex systems of formal logic it is more difficult to describe what the object can perform than to produce the object itself. Systems of high complexity possess, in principle, the ability to produce something more complex than they are themselves.

A strict definition of complexity that allows one to describe this concept quantitatively is given in the works of Kolmogorov, Chaitin, and Martin-Löf. A complex object can be coded by a certain message, say, by a sequence of zeros and units. The complexity is defined as the number of such binary digits containing all the information about the object, sufficient for its reproduction (decoding). In other words, complexity is the length of the most concise program expressed in bits which gives rise to a sequence of binary digits that describes the object.

Consider some examples. Suppose that we have two messages:

0101010101010101

0100100011011010

Which of these is more complex? Of course, the second one. The program generating the first message states $(01)^8$ or "write 01 eight times". The second message has no such abbreviated program; the

program states: "write 0100100011011010". The program has the same length as the message itself.

Let us take another example. The number 3.1416, etc., is reported. The concise program is simply the letter π which expresses the ratio of the length of a circle to its diameter.

It is remarkable that the complexity thus defined is equivalent to randomness—a random sequence of elements, say, binary digits, is characterized by complexity coinciding approximately with the length of this sequence expressed in bits. Any minimal program is of necessity random, devoid of regularity. Let us consider the program P, which is minimal for the generation of a number of binary units, a message M. If we assume that P is not random, then by definition there must exist a shorter program P′ which generates the program P. The message M can be produced by means of the following algorithm: "Calculate P from P′, then calculate M from P." In other words, in this case the program P is not minimal.

The concept of the minimal program introduced in the consideration of the concepts of complexity and randomness is, however, beset with serious difficulties. It can be easily shown that a certain sequence of binary digits is not random—it will suffice to find a program that generates this sequence and is shorter than the sequence itself. But can we state that this program is really minimal? And could we prove, on the other hand, that the given sequence is random, complex, i.e., that no shorter program exists for its reproduction?

Such statements are impossible because of the Gödel theorem of incompleteness. Modern mathematics deals with so-called formal systems. Gilbert has introduced, as such a system, a language composed of a finite alphabet of symbols, having a grammar, with the aid of which there are formulated meaningful statements, a finite number of axioms and rules for the derivation of theorems from the axioms and other theorems. In 1931 Gödel showed that any formal system of this kind cannot include all true theorems and is therefore incomplete. Gödel's proof is associated with the paradox of the Cretan Epimenides: "All Cretans are liars". Or, expressed in a different form: "This statement is false" is a true statement only if it is false. Gödel replaced the concept of truth by the concept of provability: "This statement cannot be proved." Thus, either the falsity is provable, which is forbidden, or a true statement cannot be proved and, hence, the formal system is incomplete. In a formal system it cannot be proved that a certain sequence of binary digits has a complexity larger than the number of bits in the program used to find that sequence. Consequently, finding a minimal program is always conditional to some extent and the concept of complexity is relative.

In connection with what has been said above, it should be noted that the task of science is to find a minimal program that would generate (explain) the complexity of a set of facts under study. This

is what constitutes the meaning of the "Okkam razor" in science—the essences should not be multiplied without necessity. For example, Newton's law of gravity programs the falling of an apple and the motion of planets. But the Gödel theorem does not allow one to prove the minimum, i.e., the maximum economy, of the program logically. It is exactly for this reason that logical reasonings are insufficient for the development of science. The scientific cognition requires intuition. As Mandelshtam has said, the Schrödinger equation was guessed rather than derived.

Let us return to biology. The complexity, the definition of which we have adopted, increases, as a rule, during evolution, during phylogenesis, but there are opposite situations. For example, the evolutionary transition to a parasitic way of life means a simplification rather than a complication.

As we have seen, biological systems perceive, pick out from the noise, create, reproduce and memorize information, primarily genetic information. These processes are all nonequilibrium and irreversible. The picked-out, created or perceived information is not lost, as a rule, in the subsequent biological development—during phylogenesis and ontogenesis.

The contemporary synthetic theory of evolution combines Darwin's theory, in which the key role is played by natural selection, and genetics, including its latest achievements based on molecular biology. Timofeev-Ressovsky has pointed out that in theoretical biology one feels the need for a third principle in addition to the two principles that have been established—natural selection and molecular reduplication of genes. As we have seen, this is the covariant replication, i.e., a process in which the mutational changes of DNA are fixed and stored.

The sought-for third principle is probably required for a clear-cut formulation of the factors that determine the directionality, irreversibility and, hence, the rate of evolution. The evolutionary complexity has been proposed as such a principle—the direction of evolution is determined by the direction in which complexity increases and not fitness, adaptation, and adaptability. In biological literature there is a statement that this principle is independent of natural selection. In actual fact, this is probably not so—complexity is determined by the natural selection of genotypes, i.e., by adaptations to newer and newer ecological niches. A rapid rise in complexity is provided in evolution by sexual reproduction, which leads to recombinations of genomes, to the generation of new information. The choice of a partner for reproduction is largely random if we digress from sexual selection; the emergence of progeny means the memorizing of this choice. The highest complexity of a biological object arises at the level of individuals. The evolutionary complexity (and also simplification; see below) resulting from the adaptation

to the various ecological niches means individualization. It is exactly in this sense that speciation implies a decrease in symmetry. As has been pointed out (see Sec. 15.6), there is a similarity between speciation and phase transitions.

The most complex objects in nature are individual living organisms, and among them man. Each human being is unique and cannot be coded by an abbreviated concise program. In this sense, "no substitutes exist". Obviously, what has been said also refers to the creative works of man, say, works of literature and art. It is impossible to work out a minimum program for "War and Peace"—no algorithm can be worked out for the simplification of a true work of fiction without loss of its essence.

Each organism, however, is not only an individual. Leo Tolstoy is not only a great writer, he is also a *Homo sapiens*. In other words, he is a member of the animal kingdom, of the phylum *Chordata*, the vertebrate subphylum, of the class of mammals, of the order primates, of the superfamily of anthropoids, of the family *Hominidae*, of the *Homo* genus, and of the species *Homo sapiens*. Within each of the subdivisions all its members are equivalent and can be described by a single minimal program. Here "there are substitutes", and as a member of the species *Homo sapiens* each human being, irrespective of his skin colour and other inherited characters, is equivalent to any other human being.

It is obvious that the minimal program is lengthened and the complexity increases in the hierarchy of living creatures from kingdom to species to individual. The discovery of this actually existing hierarchy in nature begun by Linnaeus is one of the stupendous events in the history of science. As has been demonstrated by the present-day evolutionary theory, the emergence of this hierarchy, which terminates in speciation, is determined by natural selection and laws of genetics.

Science, which always minimizes the program, proceeds from a species and further on. Anthropology classifies races and nationalities of humans; physiology and psychology classify the manifestations of genotypes—the structure of the body and temperaments (choleric, sanguine, melancholic, phlegmatic).

We see that the concept of complexity is relative—it depends on the level of consideration, the level of reception. At the kingdom level a member of *Homo sapiens* is equivalent to a scorpion; at the phylum level to a lancelet, at the class level to a didelphus, at the order level to a lemur, at the superfamily level to a chimpanzee, at the family level to an Australopithecus, at the genus level to a *Homo erectus*. The complexity increases with increasing similarity between the animals concerned.

For the biologist the bovine brain is a very complex system, requiring hundreds and thousands of bits for its description, but for

the butcher the description of the same brain requires no more than five bits, since the brain is only one of the nearly thirty various parts of the bull's body used for food.

We see that the complexity concept is similar to the concept of the information value considered in the preceding section. We have defined the information value as the irreplaceability, nonredundancy at a given level of consideration, the level of reception. The definition of complexity given here is perhaps similar to that of the information value—what cannot be substituted by a shorter program is indispensable, nonredundant. There are, however, differences between the two concepts.

First, the complexity concept refers to an object, to a sequence of binary digits as a whole. In contrast, one may speak of the valuability—irreplaceability, nonredundancy—of an individual element of such a sequence. We have already given the tentative determinations of the value of codons and amino-acid residues.

Second, the complexity concept characterizes only the structure of an object. On the contrary the value, i.e., irreplaceability, reflects simultaneously the functionality of the object and of its individual elements.

Thus, the concept of the information value introduced in the preceding section is broader than the complexity concept and, as a matter of fact, includes complexity. Using the notion of the value, one can overcome the difficulty associated with the possible occurrence of simplification rather than of complication during evolution. Such facts contradict the increase of complexity in biological development. One may ask: What happens in these situations with the information value?

Let us consider, as an example, the evolutionary simplification of the worm *Bonellia viridis*. The female of this animal has a macroscopic size; it is a complex multicellular organism with diverse functions. The male, on the contrary, is microscopic, lives in the sexual ducts of the female and is capable of fertilization only. Obviously, the minimal program and complexity of the male are sharply reduced as compared to those of the female. But what does it mean?

Evidently, the genome of the male does not strongly differ from that of the female. However, a considerable part of the genes of the male are blocked and do not function. The shortened program characterizes not the reduction in size of the genome but the restriction on its functionality as a result of the blocking indicated.

The program may be reduced by two factors. First, the amount of functional information may be reduced, i.e., the operating rather than the blocked genome may be reduced. Second, the text of this functional message may change and become more ordered, i.e., the functional information may become partially redundant. In this case the text is simplified and becomes less random. In the first case,

the value of each element of the message, of each letter, remains to be high and may even increase and in the second, both the complexity and the value are reduced. There is every reason to believe that in the evolutionary simplification, which is determined by the specific adaptation of the parasite or male of the worm *Bonellia* it is the first case that is realized—the complexity decreases, and the specific value of functional genetic information is not reduced and may even increase. A high value implies high adaptation (and adaptability), the irreplaceability of both structure and function.

We come to the conclusion that the third fundamental principle that may be introduced into theoretical biology is not the principle of the increase of complexity but the principle of the increase of the value, i.e., the irreplaceability of information both during phylogenetic and ontogenetic biological development. This principle is not independent of natural selection and covariant replication (i.e., of the genotype memory). However, it is exactly the formulation of this principle that stresses the irreversibility, i.e., the directionality of biological evolution.

These reasonings are, however, of the very general character. To be more specific about the principle of the increase of the value (or complexity), it is necessary to study the structure and functionality of genomes. Work in this direction has only begun.

It should be emphasized that the conventional concept of gradual evolutionary changes is insufficient because of the low probability of point mutations. If one proceeds from this concept, it will be very difficult to account for the emergence of the present-day enormous diversity of species and of the very highly complicated structure of their organs, say, the vertebrate eye, during the course of evolution.

Contemporary molecular genetics shows that a genome is a system far from being constant (Khesin). There exist factors that are transferred along the genome—plasmids, certain phages, so-called transposons. Animals have been found to display a multiplication of genes, which provides an increase in protein mass. Such multiplication was known to take place only in plants. Today it is known to occur also in prokaryotes. Obviously, the duplication of genes is the main pathway for the initial divergence of proteins, since when one gene is changed, its substitute can continue to perform the old function. As has been said above, there exist microsymbionts, mobile elements of the genome, which determine the so-called supermutability. These are some kind of infectious processes which cause mutagenic bursts. These include carcinogenesis, transduction, and integration of plasmids. In this sense, viruses and phages may be called "crazy genes".

On the other hand, the changes of genomes serve as the factors of the regulation of gene activity. Genetics is directly tied up with the biology of development. This determines such phenomena as, for

example, genetic recombinations in somatic cells—the so-called mitotic crossing-over. Intrachromosomal recombinations are possibly responsible for the generation of antibodies, immunoglobulins (see Sec. 4.8). Immunoglobulins do not obey the "one gene-one polypeptide chain" principle. Each such chain of an immunoglobulin is formed as a result of the mixing of the synthetic potentials of different genes. As a result, a nearly unlimited diversity of antibodies arise from a limited number of functional genes.

Other facts that point to changes of genomes and the regulatory role of these changes have been discovered in the study of mating behaviour in yeasts, whose cells are characterized by "sex". There has been studied the mechanism of the inverse switch of gene activity in prokaryotes—for example, a phase shift, i.e., the ability to synthesize one of the two flagellins, has been observed in Salmonella.

The gene variations that are regulatory are observed for genes responsible for the synthesis of ribosomal RNA and histones. In particular, there arise exochromosomal copies of ribosomal DNA. Presumably, all multiple, sequentially repeated genes have the ability to multiply their number.

Naturally, the changes of a genome based on recombinations are in turn subject to regulation. For example, the changes of the local structure of chromatin exert an influence on recombinations needed for the rearrangement of the genome. During ontogenesis, not only the blocking of genes occurs but also the rearrangements in the genome, in DNA, which activate or inactivate certain genes that affect the general activity of entire groups of genes.

One of the key problems of modern theory of evolution is as follows: Do the mechanisms responsible for microevolution account for macroevolution? Microevolution is defined as changes taking place within a population and governed by a shift in gene frequencies. In turn, such shifts are brought about by the natural selection of minor changes—point mutations. Macroevolution involves changes above the species level, including speciation and higher taxons (it is the evolution of major groups such as genera, families, orders, classes, and phyla).

Whereas it was believed earlier that evolution occurs slowly, by way of gradual accumulation of small changes, at present biology takes into account a multitude of facts indicating that macroevolution occurred in a jumpwise manner and was not reduced to microevolution. The absence of transient forms in the paleontological records points, in a number of cases, not to a deficiency but to the absence of such forms. Small changes are often not accumulated at all.

Vrba pointed to differences in the evolutionary behaviour of "generalists" and "specialists". The former species, which occupy wider ecological niches, are more intolerable to related species.

"Generalists" change their species features slowly, being more stable. In contrast, specialized species occupy narrow niches and are more tolerable to related species. But small changes in the environment can easily knock out any such species from the occupied niche. Therefore, "specialists" are changed and die out more rapidly, undergoing evolution at a faster rate.

Today, it is obvious that evolution has gone through numerous diverse pathways. The most important role in evolution is played by the fundamental limitations of morphological potentialities imposed by the properties of materials that form organisms at all levels of their structure. Not any organisms are possible. Why do land vertebrates have four legs? It is because their ancestors, fish, had four such limbs. No gradual transition can take place between feathers and hair, etc.

Evolution consists of the hierarchy of processes and restrictions that connect possible genotypes with actual phenotypes. The instructions coded in genomes are "filtered" through a network of developmental constraints producing a set of possible phenotypes. At this stage there operates natural selection, which preserves only those phenotypes that are adapted, i.e., have a higher value.

Detailed studies of the populations of various species of *Drosophila* have recently shown that speciation in some cases occurs as a result of the accumulation of small changes and in others, by way of "genetic revolution". In the former case, macroevolution is reduced to microevolution, and in the latter, it is not.

In the final run, the diversity of evolutionary pathways is determined by the dynamics of genes described briefly above, which is not reduced to point mutations.

We deliberately dwell here on some purely biological problems in order to emphasize the complexity and diversity of modern theory of evolution. The entire course of development of evolution theory is directly associated with Darwinism and genetics. The principle of the increase of irreplaceability introduced here, which mostly leads to the increase of complexity, is in full accord with the specific features of evolutionary processes described. The most important feature is the directionality of evolution.

Further investigations, in which evolutionary biology has to be combined with biophysics, require a detailed study of genomes. We still do not know which part of the genome is responsible for the species features and which for individual characters. Based on molecular-genetic data, which are still scarce and the available approaches, the future science will construct a model physico-mathematical theory of evolution. This theory will provide answers in quantitative form to a number of questions pertaining to the materials and rates of evolution. We are only beginning to build theoretical biology, the foundations of which were laid by Darwin and Mendel.

Recommended Literature

Chapter 1

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapter 1.
2. Volkenshtein, M. V., *Physics and Biology*, Academic Press, New York, 1982.
3. Schrödinger, *What is Life?*, Macmillan, New York, 1945.
4. Villee, G. and V. G. Dethier, *Biological Principles and Processes*, W. B. Saunders and Co., Philadelphia, 1971.

Chapter 2

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapter 2.
2. Lehninger, A. L., *Biochemistry*, Worth Publishers, Inc., New York, 1970.
3. Villee, G. and V. G. Dethier, *Biological Principles and Processes*, W. B. Saunders and Co., Philadelphia, 1971.
4. Hochachka, P. and G. Somero, *Strategies of Biochemical Adaptation*, W. B. Saunders Co., Philadelphia, L., Toronto, 1973.

Chapter 3

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapter 3.
2. Volkenshtein, M. V., *Configurational Statistics of Polymeric Chains*, Interscience, New York, 1963.
3. Birshtein, T. M. and O. B. Ptitsyn, *Conformations of Macromolecules*, Interscience, New York, 1966.
4. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Wiley-Interscience, New York, 2nd edition, 1975.
5. Tsvetkov, V. N., V. Ye. Eskin, and S. Ya. Frenkel, *Structure of Macromolecules in Solution*, Butterworth, London, 1970.
6. Bresler, S. E. and B. L. Erusalimsky, *Physics and Chemistry of Macromolecules*, Nauka, Moscow, 1963 (in Russian).
7. Dashevsky, V. G., *Conformations of Organic Molecules*, Khimiya, Moscow, 1974 (in Russian).
8. Tanford, C., *Physical Chemistry of Macromolecules*, Wiley, New York, 1961.

Chapter 4

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapter 4.
2. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 1 (in Russian); Academic Press, New York (in press).

3. Birshtein, T. M. and O. B. Ptitsyn, *Conformations of Macromolecules*, Interscience, New York, 1966.
4. Flory, P. J., *Statistical Mechanics of Chain Molecules*, Wiley-Interscience, New York, 2nd edition, 1975.
5. Pimentel, G. C. and A. L. McClellan, *The Hydrogen Bond*, W. H. Freeman, 1960.
6. Timasheff, S. N. and G. D. Fasman, eds., *Structure and Stability of Biological Macromolecules*, Dekker, N. Y., 1969.
7. Alexandrov, V. Ya., *Cells, Molecules, and Temperature*, Springer, 1979.

Chapter 5

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapters 5 and 7.
2. Vainshtein, B. K, *Diffraction of X-Rays by Chain Molecules*, Elsevier, Amsterdam, 1966.
3. Blandell, T. L. and L. N. Johnson, *Protein Crystallography*, Academic Press, London, 1976.
4. Timasheff, S. N. and G. D. Fasman, eds., *Structure and Stability of Biological Macromolecules*, Dekker, N. Y., 1969.
5. Kauzmann, W., *Quantum Chemistry. An Introduction*, Academic Press, New York, 1957.
6. Ingram, D. J. E., *Biological and Biochemical Applications of Electron Spin Resonance*, Adam Hilger, London, 1969.
7. Likhtenstein, G. I., *Spin Label Technique in Molecular Biology*, Nauka, Moscow, 1974 (in Russian).

Chapter 6

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapters 6 and 7.
2. Blumenfeld, L. A., *Problems of Biological Physics*, Springer, Berlin, Heidelberg, 1981.
3. Cornish-Bowden, A., *Principles of Enzyme Kinetics*, Butterworths, London, 1976.
4. Jencks, W. P., *Catalysis in Chemistry and Enzymology*, McGraw-Hill Book Company, New York, 1969.

Chapters 7 and 8

1. Volkenshtein, M. V., *Molecular Biophysics*, Academic Press, New York, 1977, Chapters 8 and 9.
2. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 1 (in Russian); Academic Press, New York (in press).
3. Watson, J. D., *Molecular Biology of the Gene*, 3rd edition, W. A. Benjamin, California, 1975.
4. Bresler, S. E., *Molecular Biology*, Nauka, Moscow, 1973 (in Russian).
5. Spirin, A. S. and L. P. Gavrilova, *The Ribosome*, Nauka, Moscow, 1971 (in Russian).
6. Kiselev, L. L., V. G. Nikiforov *et al.*, *Molecular Foundations of Protein Biosynthesis*, Nauka, 1971 (in Russian).

Chapter 9

1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 2 (in Russian); Academic Press, New York (in press).
2. Brillouin, L., *Science and Information Theory*, Academic Press, New York, 1956.

3. Haken, H., *Synergetics*, Springer-Verlag, B., Heidelberg, N. Y., 1978.
4. Prigogine, I., *Introduction to Nonequilibrium Thermodynamics*, Wiley-Interscience, New York, 1962.
5. Glansdorf, P. and I. Prigogine, *Thermodynamic Theory of Structure, Stability, and Fluctuations*, Wiley-Interscience, New York, 1971.
6. Rubin, A. B., *Thermodynamics of Biological Processes*, Moscow State University Press (Izd. MGU), Moscow, 1976 (in Russian).
7. Blumenfeld, L. A., *Problems of Biological Physics*, Springer, B., Heidelberg, N. Y., 1981.

Chapter 10

1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 3 (in Russian); Academic Press, New York (in press).
2. Bergelson, L. D., *Biological Membranes*, Nauka, Moscow, 1975 (in Russian).
3. Policard, A., *Cytologie moléculaire des membranes et compartiments de la cellule animale*, Masson, Paris, 1970 (Molecular Cytology of Membrane Systems of the Living Cell).
4. Vinnikov, Ya. A., *Cytological and Molecular Foundations of Reception*, Nauka, Leningrad, 1971 (in Russian).
5. Parsons, D. S., ed., *Biological Membranes*, Oxford University Press, London, 1975.
6. Vereninov, A. A., *Transport of Ions Through Cell Membranes*, Nauka, Leningrad, 1978 (in Russian).
7. Markin, V. S. and Yu. A. Chizmadzhev, *Induced Transport of Ions*, Nauka, Moscow, 1974 (in Russian).
8. Ovchinnikov, Yu. A., A. M. Shkrob, and V. T. Ivanov, *Membrane-Active Complexones*, Nauka, Moscow, 1974 (in Russian).

Chapter 11

1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 5 (in Russian); Academic Press, New York (in press).
2. Katz, B., *Nerve, Muscle and Synapse*, McGraw-Hill, London, 1966.
3. Hodgkin, A. L., *The Conduction of the Nervous Impulse*, University Press, Liverpool, 1964.
4. Khodorov, B. I., *General Biology of Excitable Membranes*, Nauka, Moscow, 1975 (in Russian).
5. Khodorov, B. I., *The Problems of Excitability*, Meditsina, Leningrad, 1969 (in Russian).
6. Tasaki, I., *Nerve Excitation. A Macromolecular Approach*, C. Thomas, Springfield, Illinois, 1968.
7. Markin, V. S., V. F. Pastushenko, and Yu. A. Chizmadzhev, *The Theory of Excitable Media*, Nauka, Moscow, 1981 (in Russian).

Chapter 12

1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 5 (in Russian); Academic Press, New York (in press).
2. Bendall, J. R., *Muscles, Molecules and Movement. An Essay in the Contraction of Muscles*, Heinemann, Ltd., London, 1969.
3. Katz, B., *Nerve, Muscle and Synapse*, McGraw-Hill, London, 1966.
4. Deshcherevsky, V. I., *Mathematical Models of Muscle Contraction*, Nauka, Moscow, 1977 (in Russian).
5. Alexander, R. M., *Animal Mechanics*, Sidgewick and Jackson, London, 1968.

Chapter 13
1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 6 (in Russian); Academic Press, New York (in press).
2. Blumenfeld, L. A., *Problems of Biological Physics*, Springer, B., Heidelberg, N. Y., 1981.
3. Lehninger, A. L., *Biochemistry*, Worth Publishers, Inc., New York, 1971.
4. Skulachev, V. P., *Transformation of Energy in Biomembranes*, Nauka, Moscow, 1972 (in Russian).
5. Skulachev, V. P., *Accumulation of Energy in the Cell*, Nauka, Moscow, 1969 (in Russian).

Chapter 14
1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 7 (in Russian); Academic Press, New York (in press).
2. The Biophysics of Photosynthesis. A collection of articles, edited by A. B. Rubin, Moscow State University Press (Izd. MGU), Moscow, 1975 (in Russian).
3. Chernavskaya, N. M. and D. S. Chernavsky, *Tunnelling Transport of Electrons in Photosynthesis*, Moscow State University Press (Izd. MGU), Moscow, 1977 (in Russian).
4. Blumenfeld, L. A., *Problems of Biological Physics*, Springer, B., Heidelberg, N. Y., 1981.
5. Konev, S. V. and I. A. Bolotovsky, *Photobiology*, Byelorussian State University Press (Izd. BGU), Minsk, 1979 (in Russian).

Chapters 15 and 16
1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 8 (in Russian); Academic Press, New York (in press).
2. Glansdorf, P. and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley-Interscience, New York, 1971.
3. Nicolis, G. and I. Prigogine, *Self-Organization in Nonequilibrium Systems*, Wiley, New York, 1977.
4. Andronov, A. A., A. A. Vitt, and S. E. Khaikin, *Theory of Vibrations*, Fizmatgiz, Moscow, 1959 (in Russian).
5. Romanovsky, Yu. M., N. V. Stepanova, and D. S. Chernavsky, *Mathematical Modelling in Biophysics*, Nauka, Moscow, 1975 (in Russian).
6. Ivanitsky, G. R., V. I. Krinsky, and E. E. Selkov, *Mathematical Biophysics of the Cell*, Nauka, Moscow, 1978 (in Russian).
7. Zhabotinsky, A. M., *Concentrational Auto-Oscillations*, Nauka, Moscow, 1974 (in Russian).
8. Hacken, H., *Synergetics*, Springer-Verlag, B., Heidelberg, N. Y., 1978.
9. Ebeling, W., *Strukturbildung bei irreversiblen Prozessen*, Teubner, Leipzig, 1976.
10. Volterra, V., *Théorie mathématique de la Lutte pour la Vie*, Gauthier-Villars, Paris, 1931.
11. Rubin, A. B., N. F. Pytieva, and G. Yu. Riznichenko, *The Kinetics of Biological Processes*, Moscow State University Press (Izd. MGU), Moscow, 1977 (in Russian).
12. Markin, V. S., V. F. Pastushenko, and Yu. A. Chizmadzhev, *The Theory of Excitable Media*, Nauka, Moscow, 1981 (in Russian).

Chapter 17
1. Volkenshtein, M. V., *General Biophysics*, Nauka, Moscow, 1978, Chapter 9 (in Russian); Academic Press, New York (in press).

2. Nicolis, G. and I. Prigogine, *Self-Organization in Nonequilibrium Systems*, Wiley, New York, 1977.
3. Oparin, A. I., *The Origin of Life on the Earth*, Academic Press, New York, 1957.
4. Bernal, J. D., *The Origin of Life*, World Publishing Company, Cleveland, Ohio, 1967.
5. Calvin, M., *Chemical Evolution: Molecular Evolution towards the Origin of Living Systems on the Earth and Elsewhere*, Oxford University Press, Oxford, 1969.
6. Eigen, M., *Selforganization of Matter and the Evolution of Biological Macromolecules*, Springer, Berlin, Heidelberg, N. Y., 1971.
7. Eigen, M. and R. Winkler, *Das Spiel*, Piper, München, 1975.
8. Eigen, M and P. Schüster, *The Hypercycle. A Principle of Natural Self-Organization*, Springer, Berlin, 1979.
9. Ebeling, W., *Strukturbildung bei irreversiblen Prozessen*, Teubner, Leipzig, 1976.
10. Romanovsky, Yu. M., N. V. Stepanova, and D. S. Chernavsky, *Mathematical Modelling in Biophysics*, Nauka, Moscow, 1975 (in Russian).
11. Schmalhausen, I. I., *Cybernetic Problems in Biology*, Nauka, Novosibirsk, 1968 (in Russian).
12. Schmalhausen, I. I., *Evolutionary Factors*, Nauka, Moscow, 1968 (in Russian); Schmalhausen, I. I., Factors of Evolution: The Theory of Stabilizing Selection, Blakiston, Philadelphia, 1949.
13. Schmalhausen, I. I., *Problems of Darwinism*, Nauka, Moscow, 1969 (in Russian).
14. Quastler, H., *The Emergence of Biological Organization*, Yale Univ. Press, New Haven, London, 1964.
15. Villee, G. and V. G. Dethier, *Biological Principles and Processes*, W. B. Saunders and Co., Philadelphia, 1971.
16. Timofeev-Ressovsky, N. V., N. N. Vorontsov, and A. V. Yablokov, *A Brief Essay in the Theory of Evolution*, Nauka, Moscow, 1977 (in Russian).
17. Grant, V., *Organismic Evolution*, Freeman, San Francisco, 1977.
18. Ebeling, W. and R. Feistel, *Physik der Selbstorganisation und Evolution*, Akademie-Verlag, Berlin, 1982.

Name Index

Number in italics refer to pages on which the selected bibliography is given.

Alexander, P., *621*
Alexandrov, V. Ya., 131, 346, *620*
Amoore, J. E., 373
Andreeva, N. S., 146
Andronov, A. A., 512, *622*
Anfinsen, C. B., 119
Arago, D., 162
Armstrong, C., 401
Arrhenius, S., 214
Astbury, W. T., 139
Avery, O. T., 236

Bachinsky, A. G., 606
Baev, A. A., 286
Baltimore, D., 280
Barnes, A., 489
Bauer, E. S., 19
Belintsev, B. N., 532
Belitser, V. A., 443
Belousov, B. P., 539
Belozersky, A. N., 42
Bendall, J. R., *621*
Berestovsky, G. N., 401
Bergelson, L. D., 349, *621*
Bernal, J. D., 145, 146, *623*
Bernstein, J., 377
Bertalanffy, L., 19, 329
Bezanilla, F., 401
Biltonen, R., 213
Birshtein, T. M., 114, *619*, *620*
Blandell, T. L., *620*
Blumenfeld, L. A., 212, 318, 456, 458, 459, 484, *620*, *621*, *622*
Bohr, N., 13, 14
Bolotovsky, I. A., *622*
Borisov, A. Yu., 483
Borisov, V. V., 146
Bragg, W., Jr., 145
Braunstein, A. E., 200, 204
Bresler, S. E., *619*, *620*
Brillouin, L., *620*
Brönsted, J. N., 29
Butenandt, A. F. J., 372
Bütner, E. K., 565

Calvin, M., 476, *623*
Chaikovsky, G., 531
Chaitin, G., 611
Chance, B., 445, 459, 547
Changeux, J. P., 219, 222, 354
Chernavskaya, N. M., 460, 498, 500, 583, *622*
Chernavsky, D. S., 212, 459, 460, 484, 498, 500, 551, 583, *622*, *623*
Chernigovsky, V. N., 436
Cherny, D., 489
Chizmadzhev, Yu. A., 369, 393, 394, *621*, *622*
Cohen, L., 401
Cone, R., 497
Corey, R. B., 34, 97
Cornish-Bowden, A., *620*
Cotton, A., 167
Crick, F. H. C., 138, 239, 242, 276, 299, 309
Crothers, D. M., 266

Danielli, J. F., 348
Darwin, Ch., 564, 585, 586, 602, 618
Dashevsky, V. G., *619*
Davson, H., 348
Davydov, A. S., 155, 423
Debye, P., 90
Delbrück, M., 15
De Robertis, E., 403, 404
Deshcherevsky, V. I., 425, *621*
Dethier, V. G., *619*, *623*
Dibrov, B. F., 598
Dickerson, R., 463
Dogonadze, R. R., 212, 213, 463, 465
Doty, P., 109, 250, 264
Duysens, L., 473

Ebeling, W., 531, 584, *622*, *623*
Edelman, G. M., 133, 376
Edman, K., 420
Ehrenfest, P., 573
Ehrenfest, T., 573
Eigen, M., 224, 288, 307, 568, 574, 576, 578, 579, *623*
Einstein, A., 497
Elsasser, W. M., 15
Elyashevich, M. A., 179
Engelhardt, V. A., 410, 411, 443, 494
Epimenides, 612
Ermishkin, L. N., 401
Erusalimsky, B. L., *619*
Esipova, N. G., 146
Eskin, V. Ye., *619*
Evdokimov, Yu. M., 262
Eyring, H., 189, 427, 459

Fabre, J., 372
Faraday, M., 174

Fasman, G. D., *620*
Feistel, R., 584
Feofilov, P. P., 158
Fesenko, E. E., 375
Finkelshtein, A. V., 121
Fischer, E., 206
Fisher, H. F., 117
Florentiev, V. L., 285
Flory, P. J., 83, 91, 103, *619*, *620*
Fong, P., 261
Fowler, R., 78
Fox, S., 567
Frank, G. M., 433
Frank-Kamenetsky, M. D., 252, 254
Franklin, R., 239
Frenkel, S. Ya., *619*
Frenkel, Ya. I., 155
Fresnel, A. J., 162

Gamow, G. A., 275
Gavrilova, L. P., 294, *620*
Georgiev, G. P., 310, 311
Gilbert, D., 612
Glansdorf, P., *621*, *622*
Gödel, K., 612
Goldshtein, B. N., 223
Gottikh, B. P., 307
Grant, V., *623*
Green, D., 457
Gribov, L. A., 179
Grossberg, A. Yu., 84
Gurdon, J., 589
Gurfinkel, V. S., 439
Gursky, G. V., 266, 267, 307
Gurvich, A. E., 161, 591

Haken, H., *621*, *622*
Haldane, J. B. S., 565
Hanson, J., 412
Hartline, H. K., 378
Hecht, S., 489
Helmholtz, H. von, 12, 13, 377, 487
Hess, H., 547
Hill, A. V., 418, 419, 420
Hille, B., 401
Hochachka, P., *619*
Hodgkin, A. L., 360, 377, 389, 390, 395, 400, 401
Hodgkin, D. C., 145, 146
Holley, R., 286
Hooker, J., 564
Huxley, A. F., 377, 389, 390, 395, 400, 401, 424
Huxley, H., 412, 416

Ičas, M., 298
Ingram, D. J. E., *620*

Ivanitsky, G. R., 557, 559, *622*
Ivanov, V. I., 369, *621*
Ivanov, V. T., 202, 285

Jacob, F., 306, 595
Jencks, W. P., 620
Ji, S., 457
Johnson, L. N., *620*
Johnston, J. W., Jr., 373
Jortner, J., 463, 464
Joule, J. P., 13

Karpeisky, M. Ya., 202
Katchalsky, A., 336, 345, 409
Katz, B., 360, 377, *621*
Kauzmann, W., *620*
Kendrew, J. C., 125, 145
Keynes, R. D., 401
Khaikin, S. E., 622
Khesin, R. B., 308, 616
Khodorov, B. I., *621*
Khokhlov, A. R., 84
Khorana, H. G., 43, 296
Kimura, M., 124
Kiselev, L. L., *620*
Kitaigorodsky, A. I., 68
Klug, A., 309
Kock, B., 473
Kohlrausch, K., 70
Kolmogorov, A. N., 611
Koltsov, N. K., 280
Kompaneyets, A. S., 392
Konev, S. V., 622
Kopun, M., 302
Kornberg, A., 268, 270
Koshland, D. E., Jr., 206, 207, 221
Kounosu, S., 430
Kramers, H. A., 79
Krasnovsky, A. A., 469
Krinsky, V. I., 557, 559, 562, *622*
Kuhn, H., 578, 581, 582
Kuhn, W., 73, 167, 260
Kuznetsov, A. M., 465

Landau, L. D., 66
Landsberg, G. S., 178
Landsteiner, K., 133
Langmuir, I., 116
Lavenda, B., 593
Lavoisier, A., 440
Lazurkin, Yu. S., 252, 254, 258, 263
Lehninger, A. L., 444, 449, *619*, *622*
Leonardo da Vinci, 439
Lerman, L., 267
Liberman, E. A., 454, 455

Lifshitz, I. M., 83, 84
Likhtenshtein, G. I., *620*
Lillie, R. S., 388
Lim, V. A., 120
Linderström-Lang, K., 132
Linnaeus, C., 614
Livshits, M. A., 598
Lomonosov, M. V., 487
London, F., 61
Longuet-Higgins, M., 260
Lotka, A. J., 515
Lumry, R., 197, 213
Lyapounov, A. M., 19, 511
Lyubimova, M. N., 410, 411

MacLeod, C. M., 236
Makarov, A. A., 146
Makarov, V. L., 173
Mandelstam, L. I., 178, 519, 613
Markin, V. S., 369, 388, 393, 394, 556, *621*, *622*
Marmur, J., 264
Martinez, H., 592
Martin-Löf, P., 611
Massey, V., 262
Matthaei, J. H., 294
Maxwell, J. C., 12, 487
Mayer, R., 13
Mayr, E., 585
McCarty, M., 236
McClare, C., 423
McClellan, A. L., *620*
Mendel, G., 618
Meselson, M., 268
Miller, S., 565, 566
Mitchell, P., 451, 453
Moffitt, W., 156
Molchanov, A. M., 538
Moncrieff, R. W., 373
Monod, J., 219, 306, 354, 595
Mukhin, L. M., 566, 567

Nasonov, D. N., 131
Nicolis, G., *622*, *623*
Nikiforov, V. G., *620*
Nirenberg, M. W., 294, 296, 298

Oesterhelt, D., 498
Okazaki, R., 274
Onsager, L., 341
Oosawa, F., 430
Oparin, A. I., 565, 567, 584, *623*
Oplatka, A., 409, 423
Orgel, L. E., 312
Oster, G., 336
Ostrovsky, M. A., 495
Ovchinnikov, Yu. A., 369, *621*

Parsons, D. S., *621*
Pasteur, L., 47
Pastushenko, V. F., 369, *621*, *622*
Pauling, L., 34, 97, 137
Perutz, M. F., 125, 144, 145, 209, 229, 231, 232
Phillips, D. C., 200
Pimentel, G. C., 620
Podolsky, R., 420
Poincaré, J. H., 510
Poletaev, A. I., 174
Policard, A., *621*
Pressman, D., 134, 137, 369
Prigogine, I., 19, 341, 579, *621*, *622*, *623*
Privalov, P. L., 128, 258
Ptitsyn, O. B., 120, 121, 150, *619*, *620*
Pytieva, N. F., 622

Quastler, G., 602, 623

Ramachandran, G. N., 101, 102, 138
Raman, C. V., 178
Rich, A., 138, 289
Riznichenko, G. Yu., *622*
Romanovsky, Yu. M., *622*, *623*
Rosenblueth, A., 559
Rosenfeld, L., 164
Rubin, A. B., *621*, *622*
Rubin, M., 373
Rumer, Yu. B., 299
Rutter, W., 589

Sanger, F., 286
Schlögl, F., 528
Schmalhausen, I. I., 602, 603, 623
Schrödinger, E., 13, 14, 15, 237, 243, 604, *619*
Schuster, P., 578, 579, *623*
Selkov, E. E., 547, 549, *622*
Sendov, B., 310
Shannon, C., 604
Sharonov, Yu. A., 178
Shemyakin, M. M., 369
Shkrob, A. M., *621*
Shoenberg, C., 420
Skulachev, V. P., 455, 495, *622*
Slater, E., 449
Somero, G., *619*
Spiegelman, S., 578
Spirin, A. S., 292, 293, 294, *620*
Stahl, F., 268
Stahl, G. E., 13
Stepanov, B. I., 179
Stepanova, N. V., *622*, *623*
Sternberger, L., 134

Stoeckenius, W., 498
Szent-Györgyi, A. I., 423

Tanford, C., 118, 129, *619*
Tasaki, I., 377, 401, *621*
Tawada, K., 430
Teilhard de Chardin, P., 585
Temin, H., 280
Teorell, T., 556
Terenin, A. N., 470, 566
Thom, R., 512
Timasheff, S. N., *620*
Timiryazev, K. A., 467
Timofeev-Ressovsky, N. V., 15, 613, *623*
Tinoco, I., 288
Tolstoy, L. N., 614
Tonomura, Y., 423
Träuble, H., 354
Tsanev, R., 310
Tsvetkov, V. N., *619*
Tsybakova, E. T., 443
Tumanian, V. G., 138, 288
Turing, A., 524, 592

Ulstrup, J., 465
Umbarger, H. E., 221

Vainshtein, B. K., 146, *620*
Vallee, B., 216
Varshavsky, Ya. M., 262
Vasiliev, V. D., 290
Vavilov, S. I., 158, 489
Vazina, A. A., 416, *433*
Venkstern, T. V., 494, 495
Vereninov, A. A., *621*
Verhulst, P., 519
Villee, G., *619*, *623*
Vinnikov, Ya. A., *621*
Vitt, A. A., *622*
Vogel, F., 302
Volkenshtein, M. V., 74, 179, 212, 223, 426, 532, 598, *619*, *620*, *621*
Volterra, V., 515, 516, 534, *622*
von Neumann, J., 611
Vorontsov, N. N., *623*
Vrba, E., 617

Waddington, C., 332, 536, 591
Wald, G., 490, 492
Wannier, G., 79
Watson, J. D., 239, 242, *620*
Weissmann, P., 578
Wessels, N., 589
Wiener, N., 559
Wigner, E., 15, 573
Wilkins, M. H. F., 239
Winfield, S., 463
Winkler, R., *623*
Witt, H. T., 348, 473, 481, 482
Wittmann, G., 279
Wöhler, F., 25
Wolpert, L., 592
Wyman, H. J., 219, 233, 354

Yablokov, A. V., *623*
Young, T., 487

Zaikin, A. N., 544, 545
Zasedatelev, A. S., 266
Zavoisky, E. K., 182
Zhabotinsky, A. M., 539, 544, 545, *622*
Zimm, B., 260, 262

Subject Index

Absolute reaction rate theory, 189 *et seq.*, 426
Absorption coefficient, 163
Accommodation, 485
Acetylcholine, 403-5
Acetylcholine esterase, 403, 404
Acetyl-CoA, 43, 44, 57, 441
Actin, 412, 414, 434-7
Actinin, 416
Action potential, 379, 381, 388, 391, 394, 395
Action spectrum, 470, 471
Activated-complex theory, 189
Activation energy, 188
Activator, 196
Active membrane transport, 347, 361-8
Active site, 198
Activity, 30, 31, 321
  coefficient, 30
  enzymic, 198
  optical, 169-74
    hidden, 173
    induced, 171
Actomyosin complex, 417
Adaptability, 613, 616
Adaptation, 18, 613, 616
Adaptive radiation, 588

Adenosine diphosphate (ADP), 43, 364, 365, 477
Adenosine monophosphate (AMP), 43, 365
  cyclic, 44, 45, 308, 309, 376
Adenosine triphosphate (ATP), 43, 56, 57, 361, 406, 417, 418, 423, 437, 440, 441 *et seq.*, 477
Adenylates, 43-5
Adhesion, cell, 376
ADP, *see* Adenosine diphosphate
Adrenalin, 45, 54
Aerobic systems, 56
Afferent neuron, 377
Affinity, 321, 326-34, 408
Alcoholic fermentation, 56
Aldimine, 200, 497
All-or-none law, 161, 355, 381, 404
Allosteric enzymes, 95, 221, 222, 233
  cooperativity of, 222
  quaternary structure of, 222
Allosteric inhibitor, 221, 222
Allosteric systems, 221
Allosterism, 221 *et seq.*, 233, 593
Alpha-helix, 98, 100, 110, 120
Amide group, 154, 180
Amino-acid replacements, 36, 124
Amino acids, 27-9, 46, 118, 280 *et seq.*
  as amphoteric electrolytes, 32
  electrochemical properties of, 32, 33
  functional similarity, 606
  hydrophobic nature of, 118
  polycondensation of, 27, 64
  see also Proteins
Amino-acid sequence, 275
Aminoacyl-tRNA synthetase, 281
Amoore receptor-site hypothesis, 373, 374, 375
Amorphous halo, 148
Amount of information, 22, 313 *et. seq.*
AMP, *see* Adenosine monophosphate
Amylase, 205, 207
Amylopectin, 50
Amylose, 50
Anaerobic systems, 56
Andrew cross, 147
Animal fats, 50
Anomalous ORD, 166 *et seq.*
Antibody, 132 *et seq.*, 586, 596 *et seq.*
Anticodon, 281, 287, 299
Antigen, 132 *et seq.*, 586, 596 *et seq.*
Antigen-antibody complex, 136, 137
Antiport, 453
Aperiodic crystal, 14, 96, 119, 340, 604
Apoenzyme, 51, 216
Archibald method, 88
Arrhenius formula, 22, 188, 191, 263, 371
Artificial membranes, 350
Aspartate aminotransferase, 146, 172, 200, 203, 216, 217, 221, 222
Astrophysics, 16
Asymmetric carbon atom, 45
Atom-atom potential method, 68
ATP, *see* Adenosine triphosphate
ATPase, 43, 95, 362, 364, 365, 417, 481
ATPase activity, 411, 417, 437
ATP synthetase, 450
Autocatalysis, 343, 344, 521-7, 529-31
Auto-oscillations, 432, 513, 538 *et seq.*
Autotrophs, 565
Auto-wave processes, 539 *et seq.*
Avogadro's number, 86
Axial texture, 147
Axon, 377-88
Axoplasm, 380, 381, 382

Bacterial photosynthesis, 467
Bacterial transformation, 236
Bacteriophage φ X174, 270, 271
Bacteriophages, 42, 258, 346
Bacteriorhodopsin, 497-500
Becquerel's formula, 174, 175
Beer-Lambert law, 154, 163
Benard convection, 344
Benard instability, 345
Berg nomogenesis, 585
Beta-form, 98-101, 123
Bifurcation, 512, 522
Biochemical reactions, 19, 55 *et seq.*
Biochemistry, 12, 23, 25, 539
Bioenergetics, 12, 43, 55, 346 *et seq.*, 440 *et seq.*
Biogeocenosis, 16
Bioinorganic chemistry, 23, 25, 54, 216
Biological clocks, 538
Biological evolution, 14, 18, 21, 26, 36, 55, 347, 564-8, 585 *et seq.*
Biological field, 13, 591
Biological individuality, 17
Biological macromolecules, 63 *et seq.*, *see also* Biopolymers
Biological oxidation, 440 *et seq.*
Biological processes
  dynamical, modelling of, 501 *et seq.*
  periodic, 538 *et seq.*
  photo-, 466 *et seq.*
Biology, 11, 12, 17, 539
  and chemistry, 25 *et seq.*
  and physics, 13, 14, 15

Biomechanics, 437-9
Biomolecules, 23, 26, 58, 45 *et seq.*
Bionics, 439
Bioorganic chemistry, 23, 25, 71, 168
Biophysics, defined, 11, 12
  chemical foundations of, 25-62
  methods of, 22-4
  relation to other sciences, 12, 16, 18
  scope of, 22, 23
Biopolymers, 14, 15, 23, 58, 63 *et seq.*, 103, 535
  optical activity of, 169 *et seq.*
  solutions of, 148, 149, 150
  structure of, methods of investigation, 142 *et seq.*
Biopotentials, 347, 378
Biosphere, 14, 16, 21, 22, 332
Biosynthesis, 36, 43, 45, 55, 56, 119, 275, 280 *et seq.*
Bit, 314
Black box, 536
Blast cells, 597, 598
Blastula, 590
Blumenfeld-Chernavsky hypothesis, 459 *et seq.*
Bohr effect, 229, 233
Bohr magneton, 186
Boltzmann's constant, 73, 315
Boltzmann distribution, 70, 184, 188, 189, 317
Boltzmann factor, 59, 183
Bombykol, 372
Born-Oppenheimer theorem, 179
Botany, 12, 17
Bovine ribonuclease, 36, 37
Boyle point, 84
Bragg equation, 142
Brownian motion, 137, 260, 262
Brusselator, 521-3, 543
Buffer solutions, 33, 34
Burnet clonal-selection theory, 596
Butane, 69, 74, 75

Calorimetric methods, 23
Calvin cycle, 476, 477, 551-3
Cancerogenosis (carcinogenesis), 237, 616
Carbohydrates, 46, 49, 50, 51, 55, 57, 441
Carboxyhemoglobin, 172
Carboxypeptidase, 208
Carnot cycle, 407
Carotene, 52, 153, 468
Carotenoids, 474, 479, 481, 490
Carriers, 339, 362, 369, 370, 445, 458
Catalysis, 188 *et seq.*
Catastrophes, theory of, 512
Causality, 18, 19
Cell biophysics, 22, 24
Cell differentiation, 589 *et seq.*
Cell membranes, 346 *et seq.*
Cellular immunity, 132, 596 *et seq.*
Cellular respiration, 57, 448
Cellulose, 50
Central dogma, 280
Central nervous system, 377
Centre, 508, 511
Centrifugation, 86, 87
  density gradient, 88, 89
Chain flexibility, 71, 72, 74 *et seq.*
Chain length, 82, 83
  contour, 83
  persistent, 83
  rms, 82
Channels, membrane, 352, 399-402
Chargaff's rules, 41, 42, 239, 245
Chemical evolution, 26, 27
Chemical kinetics, 188-92
Chemical microreactors, 584
Chemical relaxation, 223, 224
Chemical shift, 185
Chemical transmitter, 402
Chemiluminescence, 161
Chemiosmotic hypothesis, 450 *et seq.*, 481
Chirality, 45-9, 172, 173, 581
Chitin, 50
Chlorella, 471
Chlorophylls, 53, 57, 467 *et seq.*, 472 *et seq.*
  properties of, 474
Chloroplasts, 347, 441, 469-80
Cholesterol, 348
Chromatin, 152, 309
Chromatography, 35, 36
Chromophore, 155, 168, 172
Chromosomal mutation, 301
Chromosome, 237, 280, 301, 309, 347, 590
Chronaxy, 381, 382
Chymotrypsin, 205, 207
Chymotrypsinogen, 129
Cilia, 372, 433, 434
Ciliary muscles, 485
Circadian rhythms, 538, 539
Circular birefringence, 162, 174
Circular dichroism, 161, 166, 167, 169-73, 182, 245, 352
Clapeyron equation, 64
Clathrates, 117
Clonal-selection theory, 596
Clones, 597, 598
Clostridium acidiurici, 132
Clostridium pasteurianum, 132

Clostridium thermoaccharolyticum, 132
Clover-leaf model, of tRNA, 287, 288, 289
Coacervates, 567, 584
Code ratio, 276, 277
Codon, 278, 279, 295-301, 606, 608
Coenzymes (cofactors), 43, 51, 52, 198, 200
Coil, 73, 84, 91, 92, 96
Coil-globule transition, 84, 85
Colicins, 354
Collagen, 101, 132, 138, 139, 146, 148, 409, 439
Compartmentation, 26, 59, 346, 347, 514, 567, 578, 584
Compensation effect, 197, 198
Competitive inhibition, 196
Complementarity, 137, 214, 216
Complementarity principle, 14
Complexity, in evolution, 611-8
Concentration waves, 527, 535
Concerted general acid-base catalysis, 204
Configuration, 71
Conformational energy, 100, 101, 213, 214
Conformations, 66 *et seq.*
Conformer, 70, 74, 75
Conformon, 215, 355, 457, 461, 462
Congruent systems, 202
Constitutive mutants, 306
Contractile proteins, 95, 411, 416, 437, 439
Convergent phenomena, 588
Cooperativity
  cell membranes, 354, 355
  enzymes, 217-24, 556
  helix-coil transitions, 110, 111, 112
  macromolecules, 78-82, 410
  membrane transport, 366
  neurosecretion, 405
Coordination number, 106
Cortisone, 54, 55
Cosmology, 16
Cotton effect, 167
Coulomb interaction, 101
Coulomb's law, 59
Counterions, 91, 92
Coupled reactions, 325-9
Coupling
  chemiosmotic, 449-57
  of chemical reactions, 325-9, 335-9
  of fluxes, 322-5
Creatine kinase, 199
Creation of information, 16, 48, 572
Creode, 591
Cristae, 446, 447, 448
Cross-β-form, 98, 100
Cross-bridges, 416, 417, 428
Cross-striated muscles, 411
Curie principle, 324, 335, 361
Cyclopentanophenanthrene, 54
Cytidine triphosphate (CTP), 40, 222, 223
Cytochrome *c*, 36, 178, 463-5
Cytochrome *f*, 474
Cytochromes, 36, 53, 57, 124, 152, 178, 216, 442, 463-5, 474
Cytoethology, 346
Cytoplasm, 42
Cytric acid cycle, 57, 441 *et seq.*

Danielli-Davson model, 348, 349, 350
Darwin's theory, 12, 17, 125, 568, 574, 585, 586
Davydov splitting, 155, 156, 157
de Broglie relation, 213
de Broglie standing waves, 212
de Broglie wavelength, 151
Debye (unit), 59
Debye-Hückel reciprocal length, 31
Debye-Hückel theory, 30, 31
Debye-Scherrer diagrams, 147, 148
Degree of dissociation, 30
Degree of ionization, 32, 109, 410
Degree of polarization, 158
Degree of polymerization, 114
Deletion, 276, 277
Denaturation
  of nucleic acids, 245, 250, 270
  of proteins, 96, 117, 128-32, 145, 160
Deoxyhemoglobin, 172
Deoxyribonucleic acid (DNA), 38, 41, 42, 147
  base pairing in, 239
  chain, 41
  circular, 247, 258, 270
  conformations of, 24, 240, 241
  denaturation of, 245
  double helix of, 238, 239 *et seq.*
  unwinding of, 260 *et seq.*
  forms of, 240, 241
  geometry of, 240
  hypochromism of, 245
  interaction model of, 267
  interaction with small molecules and ions, 265-7
  intramolecular interactions in, 247-50
  melting of, 258, 259
  model of, 239
  Na-salt of, 240
  nucleotide composition of, 41, 236,

Deoxyribonucleic acid (DNA)
persistent length of, 244, 247
phage, 308
renaturation of, 264, 265
replication of, 95, 237, 239, 242, 243, 268-74, 347
secondary structure of, 239
selfish, 312
semiconservative model of, 243
synthesis of, 270 *et seq.*
tertiary structure of, 247
torus of, 247
Deoxyribose, 38, 40, 240
Dephosphorylation, 362, 363
Depsipeptides, 369
Detergents, 348
Determinant group, 134, 135, 137
Determined eye, 609
Deuterium exchange, 125, 181
Dichroism, 181, 495, 496
circular, 161, 166, 167
Differentiation, cell, 589 *et seq.*
Diffuse X-ray scattering, 148, 149, 150
Diffusion coefficient, 87, 88, 89
Dimethyl ether, 104
Dinitrophenyl hapten, 135
Dipolar ion, 27
Dipole moment, 59
Dipole strength, 155
Direct cooperativity, 218
Disorder, 14, 48
Dispersion forces, 60
Dissipation function, 320 *et seq.*, 333, 355, 356, 363, 502, 503
Dissipative structures, 341-5, 501, 502, 503
Dissociation constant, 29
Distributed systems, 514, 523 *et seq.*
Disulphide linkage, 115, 133
Divergent evolution, 588
DNA-polymerase, 269, 270, 271
Domains, 120 *et seq.*, 133, 137, 150
Donnan equilibrium, 359
Doppler effect, 152
Double helix, 239, 242
melting of, 248, 250-60
unwinding of, 260-5
Drude formula, 169, 170
Dynamic biological processes, 501 *et seq.*
Dynamic order, 340, 341, 501
Dynamics, of genome, 616
Dynamooptical effect, 90

Ebeling-Feistel model, 584
Ecdysone, 591
Ecological niche, 18, 613
Ecology, 12
Edelman hypothesis, 376
Effector cells, 597
Efferent neuron, 377, 378
Eigen theory, 568, 572, 574 *et seq.*, 583, 584
Einstein formula, 87, 89
Elastic force, 64, 65, 73
Elasticity, 64, 65
Elastic modulus, 64, 65, 436
Elastin, 439
Electrochemical potential, 357
Electrolytes, 29-34
Electromechanochemical model, 457, 458
Electron-harmonic oscillator model, 213
Electronic absorption spectra, 153-7
Electronic conductivity, 153
Electronic-conformational interactions, 131, 212-6, 229, 232-5, 365, 457-62
Electronic transitions, 153 *et seq.*
Electron microscopy, 23, 349
Electron paramagnetic resonance, 23, 186, 187, 216
Electron-transport chain, 57, 58, 441 *et seq.*, 448, 450, 458, 484
Electrophoresis, 93, 125
zone, 93
Electrostatic interactions, 33, 130, 131, 210
Electrostriction, 127
Ellipticity, 177
Embryogenesis, 18
Emerson effect, 471, 472, 473
Enantiomers, 45
Endergonic reactions, 328, 329
Endocrine gland, 54
Endoplasmic reticulum, 347
End-to-end distance, 72, 83
Energy barrier, 189
Energy migration, 160, 161
Energy of activation, 188
Energy-rich phosphate compounds, 43
Enniatin B, 370
Entelechy, 13
Enthalpy, 116, 117, 197
Entropy, 14, 64, 65, 73, 117, 197, 315, 317, 320, 329, 602
Entropy change, 20, 22, 269, 320
Entropy production, 19, 20, 320, 330, 331, 340, 341 *et seq.*
Enzymatic reactions, 45, 193 *et seq.*, 365, 366, 467, 469
kinetics of, 193-8, 205
Enzyme action, 198-205, 437

Enzyme catalysis, 192, 236
Enzymes
  action of, 198-205
  allosteric, 95, 222, 233
  conformational properties of, 205-9
  cooperative properties of, 217-24
  induced synthesis of, 306
  physics of, 188 *et seq.*
  redox, 53, 57
  turnover number of, 198
Enzyme-substrate complex, 192, 193 *et seq.*, 198 *et seq.*, 223
Enzyme-substrate interactions, 198 *et seq.*
  physics of, 209-16
Epigenetic landscape, 591
Epinephrine, 54
Equilibrium constant, 29, 30, 105, 136, 188, 325
Erythrocytes, 90, 224, 357
Escherichia coli, 55, 247, 291, 306, 354, 435
  chemical composition of, 55
Esterases, 199
Ethane, 66, 67, 68
Ethanol, 104
Eukaryotes, 237, 238, 309, 311
Evolution, 12, 15, 16, 17, 22
  and complexity, 611-8
  chemical, 26, 27
  biological, 14, 18, 21, 26, 36, 55, 538, 564 *et seq.*, 585
  Darwin's theory of, 12, 17, 125, 568, 585, 586
  macro-, 617, 618
  micro-, 617, 618
  of early life, 564 *et seq.*
  of genetic mechanism, 579
  of stars, 16, 17
  of viruses, 16, 17
  prebiotic (prebiological), 39, 48, 347, 538, 568-72
EXAFS, 151
Exciton splitting, 170
Exciton transition, 171
Exergonic reactions, 329
Extinction coefficient, 128, 154, 163, 175
Eyring formula, 191
Eyring theory, 427, 428

Facilitated transport, 339, 362
Faraday effect, 174 *et seq.*
Faraday's number, 357
Far-from-equilibrium processes, 339-45
Fenn effect, 422
Fermat's principle, 19
Ferrodoxin, 53, 132, 152
Ferroheme, 226
Ferroprotoporphyrin, 226
Fibrillation, 559-63
Fibroin, 141
Fick's equation, 89
Fine structure, 187
Flagella, 372, 433, 434
Flavin adenine dinucleotide (FAD), 52, 443
Flavin mononucleotide (FMN), 52
Flexibility, kinetic, 78
  thermodynamic, 74
Flight muscles, of insects, 432
Flory theory, 83, 84, 91
Flory θ-point, 84
Flow birefringence, 90
Fluid mosaic model, 350
Fluorescence, 158, 160
Fluxes, 320, 322-5
Fly wings, motion mechanism of, 438
Focus, 508-12
Fong theory, 262
Forces and fluxes, 320
Frameshift mutation, 277, 301
Free energy, 64, 197, 317, 321
Free-energy change, 118, 466
Free-energy profile, 189
Freely jointed chain, 72, 74
Free radicals, 187
Freeze-etching method, 349
Frictional coefficient, 86, 87, 89
Frobenius theory, 80
Frogs, thermophilicity of, 131
Fuel element, 451

Galvani experiments, 418
Gamma-resonance spectroscopy, 23, 152
Gamow "diamond code", 276
Gating current, 400, 401
Gene multiplication, 616
Generalized flux, 320
Generalized force, 320
Gene regulation, 96, 306-12
Genetic code, 236, 275 *et seq.*, 297
  deciphering of, 294 *et seq.*
  degeneracy of, 278
  universality of, 278, 279
  wobble hypothesis in, 299
Genetic engineering, 286
Genetic recombination, 586
Genetic variability, 12,17
Genome, 613, 614, 617
Geophysics, 16
Giant squid axon, 380 *et seq.*
Gibbs free energy, 317, 321

Glial cells, 377
Globular proteins, 114-9, 120, 121 *et seq.*
Globule, 84
  structure of, 125-32
Globule-coil transition, 145
Glucose, 49
  mutarotation of, 204
Glutamate dehydrogenase, 207
Glycine, 200
Glycogen, 50, 57, 58
Glycolysis, 43, 56, 57, 547-51, 565
Glycoproteins, 348, 376
Glycose, 49
Gödel theorem, 612, 613
Gramicidin-A, 371
Granum (grana), 477
Green complexes, 445
Green-Ji model, 457, 458
Gyromagnetic ratio, 183

Hairpins, 120, 122, 123
Haldane law, 235
Halobacteria, 497
Halobacterium halobium, 497
Hamilton's principle, 19
Hapten, 134, 135, 137
Harmonic oscillators, 60, 61, 214
Heavy meromyosin (HMM), 413, 414, 416, 417
Heisenberg uncertainty principle, 14
Helix-coil transitions, 108-14, 157, 249-65, 270, 411
Helmholtz free energy, 317
Helper cells, 598
Heme, 53, 225-35, 463
Heme-heme interaction, 225 *et seq.*
Hemoglobin, 36, 53, 90, 95, 115, 125, 152, 172, 178, 216, 224-35
  absorption spectrum of, 225
  Bohr effect in, 229, 233
  carbon monoxide binding by, 235
  cooperative properties of, 225, 232
  dissociation of, 233
  electronic-conformational interactions in, 229, 232
  heme groups of, 226
  Hill equation for, 228, 229, 231
  magnetic properties of, 226
  mutant, 36, 227
  oxygen affinity of, 228, 229
  oxygenation of, 228, 229, 231
  oxygen binding by, 225, 235
  quaternary structure of, 232, 235
  salt bridges in, 230, 232
  structure of, 227, 232, 235
  subunits of, 225, 232, 233
Heredity, 12, 25, 236
Heterogeneous catalysis, 192
Hexokinase, 207
Hexose monophosphate shunt, 56
Hill equation (hemoglobin), 228, 229, 235
Hill equation (muscle), 420, 421, 425, 427
Hill graph, 234
Hill reaction, 469, 471
Histones, 124, 309, 310
Hodgkin-Huxley equation, 392, 557
Hodgkin-Huxley theory, 389-92
Hodgkin-Katz formula, 361, 382
Holoenzyme, 51
Homeorhesis, 332, 536
Homeostasis, 332, 536
Homogeneous catalysis, 192
Hooke's law, 64, 82
Hormones, 54, 55, 95, 589
Humoral immunity, 132
Hydrogen bonds, 62, 97, 98, 100, 103 *et seq.*, 151
Hydrolysis, 35, 36
Hydrophobic interactions, 62, 114 *et seq.*, 121, 122, 123, 127, 128, 130
Hydroxyapatite, 438
Hydroxylysine, 138
Hydroxyproline, 138
Hyperchromism, 155, 156, 157, 245, 246, 250
Hyperfine splitting, 187
Hyperfine structure, 185, 186
Hypochromism, 156, 157

Image forces, 369
Immune memory, 132, 598
Immune response, 595 *et seq.*
Immunity, 132-7, 376, 596-602
Immunoglobulins, 132, 133, 149, 597, 617
  models of, 149
Indirect cooperativity, 219
Indirect coupling, 335-9
Induced-fit effect, 204, 206
Induced-fit theory, 206
Induced optical activity, 171
Inductive forces, 60
Informational macromolecules, 63
Information theory, 15, 22, 313 *et seq.*, 602 *et seq.*
Information value, 576, 603 *et seq.*
  selective, 570, 576
Inhibition, 196, 221, 222
Inhibitor, 196, 221
  allosteric, 221

Inorganic phosphate, 43
Insertion, 276, 277
Instability, 16, 344, 345, 502
Insulin, 54
Intercalation, 267
Internal energy, 64, 65, 407
Internal rotation, 66 *et seq.*, 100
Intramolecular interactions, 247-50
Intrinsic viscosity, 86, 91
Introns, 311
Ion-dipole interactions, 59
Ion-exchange resins, 93
Ionic atmosphere, 30
Ionic bonds, 59
Ionic channels, 399-402
Ionic permeability, 399 *et seq.*, 480
Ionic strength, 31
Ionites, 93
Ionophores, 369, 371
Ion transport, 368-72
Iris, 485
Ising model, 79, 111, 121
Isoclines, 506, 507
Isoelectric point, 32
Isoelectronic substances, 104
Isoenzymes, 586
Isomerization, 71
Isometric contraction, 418-22
Isomorphous replacements, 144
Isotonic contraction, 419

Jacob-Monod trigger circuit, 595 [illegible]

Katchalsky-Oplatka machine, 409
Keratin, 138, 139, 140, 141, 148
Kerr effect, 90
Ketimine, 200
Kimura's neutralistic theory, 124
Kinematic chains, 438
Kinetic branch, 344
Kinetic formaldehyde method, 263, 264
Kinetic theory, of rubber elasticity, 73
Kinks, 354, 399
Kitaigorodsky method, 68
Kornberg synthesis, 270, 271
Koshland theory, 206, 207
Krasnovsky reaction, 470
Krebs cycle, 57, 441 *et seq.*, 448, 577
Kronig-Kramers relations, 167
Kuhn kinetic theory, 73
Kuhn model, 167, 168
Kuhn theory, prebiotic evolution, 581, 582, 583

Lac operon, 306
Lac repressor, 307
Lamella, 477
Landau-Lifshitz theorem, 251
Langmuir isotherm, 195, 218, 220, 228, 272
Large-angle scattering, 149, 150
Larmor precession, 174, 182
Laser, 340, 527
Latent period, 562
Laue pattern, 143
Layer lines, 147
Leading centres, 544, 545, 561
Le Chatelier's principle, 19
Lecithin, 348
Leghemoglobin, 146
Lennard-Jones potential, 68, 69, 101, 102
Lenz's law, 19
Levshin-Perrin formula, 158
Life, 14
  origin of, 564 *et seq.*
Lifshitz theory, 128
Ligand field theory, 227
Light meromyosin (LMM), 413, 414
Light scattering, 23, 89, 90
Lillie model, 388
Limit cycle, 513
Lim theory, 120, 121
Linear memory, 83
Lineweaver-Burk equation, 195
Lipids, 49, 50, 51, 348 *et seq.*, 441
Liquid crystals, 351, 416, 424
Living organisms, 15, 16, 17, 19, 22, 25, 29, 319
Lock-and-key model, 206
Lohmann reaction, 418
London forces, 60
Loops, 246
Lorentz factor, 169
Lorentz-Lorenz formula, 164
Lotka-Volterra model, 515-20, 523, 543
Luminescence, 157-61
  energy transfer diagram, 159
  excitation energy, 157
  Levshin-Perrin formula, 158
  polarization, 158, 159, 160
  quenching, 159
Luminophors, 160, 161
Lymphocytes, 132, 586, 596 *et seq.*
Lymphoid cells, 597
Lysozyme, 146, 149, 150, 200, 201, 205, 208, 209, 278
  structure of, 150, 201

Macroevolution, 617, 618
Macroions, 91
Macromolecules, 14, 15, 23, 410
  and rubber elasticity, 63-6
  as a cooperative system, 78-82
  informational, 63
  methods of investigation of, 85 *et seq.*
  physics of, 63 *et seq.*
  rotational-isomeric theory of, 71-8
  self-organization of, 568 *et seq.*
Macrotetrolides, 369
Magnetic circular dichroism, 174-8, 232
Magnetic optical activity, 174-8
Magnetic rotation, 23, 174-8, 232, 233
Magnetic rotatory dispersion, 232, 233, 234
Magnetic transition dipole moment, 164
Magnetobiology, 227
Maniloff model, 271
Markov chains, 504, 588
Master kinetic equation, 532-5
Maxwell dynamooptical effect, 90
Maxwell model, 432
Mechanical filtration capacity, 356
Mechanochemical cycle, 407, 408
Mechanochemical machine, 407, 408, 409
Mechanochemical processes, 54, 406 *et seq.*, 433 *et seq.*
Mechanochemical systems, 406 *et seq.*, 433-7
Mechanoreception, 433, 436
Mediated transport, 362, 369
Meiosis, 433, 435
Membrane(s), 50, 58, 96, 346 *et seq.*, 377 *et seq.*
  bioenergetics, 346 *et seq.*
  conformational properties of, 351-5
  depolarization of, 382, 384
  dynamics in, 554-9
  excitation of, 377 *et seq.*
  selectivity of, 357
  structure of, 348-51
Membrane transport, 355 *et seq.*, 368-72
  active, 347, 361-8
  passive, 347, 355-61
Memory cells, 597, 598
Meromyosin, 413, 414, 416
Messenger RNA (mRNA), 42, 281
  breakdown of, 311
  in protein biosynthesis, 281 *et seq.*
  processing of, 311
  synthesis of, 307, 308
Metabolism, 12, 43, 55, 56, 57, 58, 95, 361, 367, 441, 568, 569
Metalloenzymes, 216
Metmyoglobin, 146
Micelles, 116
Michaelis constant, 194, 195
Michaelis-Menten equation, 194, 195
Michaelis-Menten kinetics, 217
Microcalorimetry, 125, 254, 258
Microevolution, 617, 618
Microspheres, 567
Microsymbionts, 616
Mie effect, 90
Miller indices, 143
Miraculin, 376
Mirror-image antipodes, 45
Missense mutation, 301
Mitchell chemiosmotic hypothesis, 450-7, 480, 481, 498
Mitochondria, 279, 347, 352, 440 *et seq.*
  energy production in, 440
  functions of, 440, 446, 457, 458
  properties of, 449
  structure of, 446, 447, 448, 457
Mitogenetic rays, 161
Mitosis, 433, 435
Mitotic apparatus, 435, 436
Modifier, 196
Modulus of elasticity, 64, 65, 436
Moffitt equation, 170, 171
Molar absorptivity, 154
Molar rotation, 162
Molecular adaptation, 26
Molecular biology, 14, 15, 23, 25, 26, 119, 168, 236, 237 *et seq.*
Molecular biophysics, 21, 22, 23, 24, 63, 145
Molecular ellipticity, 163
Molecular flexibility, 137
Molecular genetics, 119
Molecular interferometry, 163
Molecular reception, 372-6
Molecular recognition, 13, 119, 282, 376
Molecular signalling, 313, 376
Molecular weight, 85 *et seq.*
Monactin, 369, 370, 371
Monellin, 375
Monod-Wyman-Changeux model, 219-21, 354
Morphogenesis, 376, 544, 589 *et seq.*
Mössbauer effect, 23, 152, 225, 232
Mössbauerography, 152, 153
Motor neuron, 377, 378
Multiplication, of genes, 616

Muscle, 377, 411 *et seq.*, 559
and locomotion, 431-3
chemistry and physics of, 417-22
ciliary, 485
contraction of, 418-22
theory of, 423-31
excitability, 401, 559
fibres, 411-4
flight, 431, 432
Hill equation for, 420, 421, 425, 427
kinetic properties of, 431-3
model of, 431, 432
proteins, 412-6
structure of, 411-6
types of, 411
Mutagenesis, 248
Mutagens, 265
Mutants, 276, 303
Mutations, 276, 300, 301 *et seq.*
chromosomal, 301
frameshift, 277, 278, 301
gene, 301
induced, 301
lethal, 305
missense, 301
neutral, 586
nonsense, 301, 305
point, 301, 302, 303
reverse, 276, 277
silent, 300, 301
spontaneous, 301
suppressor, 276, 277
transitions in, 303
transversions, 303
Mycoplasma laidlawii, 15
Myelin, 377, 378, 387
Myofibrils, 411, 413
Myofilaments, 412
Myoglobin, 95, 125, 126, 129, 145-8, 149, 152, 178, 209, 216, 224-35
binding of oxygen by, 224, 225
optical rotatory dispersion of, 232
structure of, 122, 125, 126, 225, 227
Myoneural junction, 402, 403
Myosin, 138, 410, 413, 416, 434-7
ATPase activity of, 417
components of, 413
in muscles, 412 *et seq.*
interaction with actin, 417
structure of, 414
Myxomycetes, 435, 591

NAD and NADP, 43, 44
as coenzymes, 43
in anaerobic respiration, 56
in photosynthesis, 476, 477
Natural rubber, 63-6, 73
Natural selection, 124, 613
Neogloboquadrina pachyderma, 47
Nernst equation, 360
Nervous impulse, 377 *et seq.*, 388
generation of, 394-9
saltatory transmission of, 386, 387
synaptic transmission of, 55, 388, 402-5
Network thermodynamics, 345
Neural image, 486, 488
Neuromuscular junction, 402, 403
Neuron, 377, 378, 489
Neutralistic theory, 124
Neutral mutation, 124, 586
Neutron diffraction, 150, 151, 152
Newton's second law, 426
Nitrogenase, 54, 151
Nitrogeneous bases, 38, 39, 46, 248
Node, 508-12
Node of Ranvier, 378, 386, 387, 388
Nomogenesis, 585
Noncompetitive inhibition, 196
Non-Darwinian evolution, 16
Nonenzymatic translation, 294
Nonequilibrium thermodynamics, 19-22, 319 *et seq.*, 408
Nonexciton transition, 171
Nonhistone proteins, 309, 310
Nonsense mutation, 301
Nuclear magnetic moment, 183
Nuclear magnetic resonance, 23, 125, 182 *et seq.*
Nuclear magneton, 183, 186
Nuclear physics, methods of, 150
Nuclear spin number, 183
Nucleic acids, 26, 36 *et seq.*, 236
as polyelectrolytes, 91
components of, 38
conformational properties of, 237
hypochromism in, 156
physics of, 236 *et seq.*
semiconservative model of, 243
structure of, 236, 239 *et seq.*
synthesis of, 54
Nucleoproteins, 309
Nucleosides, 40
Nucleoside triphosphates, 40, 269
Nucleosomes, 309, 310
Nucleotides, 29, 40
Nucleotide sequence, 275 *et seq.*

Odorous compounds, 373
Odour, 372-5
Ohm's law, 395
Oka formula, 74, 75
Okkam razor, 613
Olfactory receptors, 374, 375

Oligopeptides, 55
Omega principle, 585
Ommatidia, 488
One-electron model, 168
Onsager theorem, 322, 328, 330, 341
Ontogenesis, 12, 15, 237, 345, 524, 589-96, 608
Oparin theory, 565, 567
Open systems, 14, 19-22, 329, 330, 331, 538
Operator gene, 306, 307
Operon, 306-8, 593, 595
Opsin, 490-3
Optical image, 486, 488
Optical rotatory dispersion, 166 *et seq.*
Optic nerve, 485, 486
Oregonator, 542
Order through fluctuations, 503, 526
Orientation effect, 204
Orientation forces, 59
Oscillator strength, 154, 155, 157
Osmometry, 85
Osmotic flow coefficient, 356
Osmotic pressure, 85
Oxidation-reduction potential, 443, 466
Oxidation-reduction processes, 51, 56
Oxidative phosphorylation, 43, 56, 57, 95, 347, 443 *et seq.*
Oxyhemoglobin, 172
Oxytocin, 54

Pacinian body, 436
Palindromes, 311
Paracrystal, 148
Partition function, 79, 80, 81, 190
Passive membrane transport, 347, 355
  Donnan equilibrium, 359
  mechanical filtration capacity, 356
  reflection coefficient, 357
Penicillin, 145
Pentose-phosphate pathway, 56
Pepsin, 146
Peptide bond, 27, 34, 35, 97, 98
Peptides, 36, 97
Perfusion experiments, 380, 381, 387
Peripheral nervous system, 377
Permeability, ionic, 358 *et seq.*, 399
  coefficient, 360, 361
Perpetuum mobile, 13
Persistent chain length, 83, 244, 247
Phage RNA, 246
Phase portrait, 505 *et seq.*
Phase-trajectory equation, 506
Phase transitions, 79, 84, 85, 128, 340, 353, 355, 559, 589, 595
Phenomenological coefficients, 322
Phermones, 372
Phosphate-oxygen ratio, 443, 444
Phosphogluconate pathway, 56
Phospholipids, 50, 51, 348, 350
Photodichroism, 495, 496
Photodissociation, 565
Photogenerator, 495
Photopigments, 490
Photosensor, 495
Photosynthesis, 24, 43, 53, 57, 347, 348, 440, 466-84, 565
  action spectrum, 470, 471
  Emerson effect, 471, 472, 473
  energy balance, 466, 467
  energy conversion, 470
  enzymatic reactions, 467, 468, 469
  free-energy change, 466
  Krasnovsky reaction, 470
  mechanism, 481-4
  oscillations, 551-4
  photosystems, 472 *et seq.*
  pigments, 467, 468, 470, 471
  quantum yield, 471, 473
  rate of, 467
  red drop, 471, 472
Photosynthetic equation, 466
Photosynthetic units, 467, 469
Photosystems, 472 *et seq.*
Phycoerythrin, 87
Phylogenesis, 12, 15, 345, 585, 608
Physics, 11, 12
  and biology, 13-5
Physiology, 12, 539, 559, 614
Pigments
  photosynthesis, 468
  visual, 490
Planck's constant, 61
Plasmids, 286
Plasmodium, 435
Plastoquinones, 474, 475
Point mutation, 301, 303
Poisson's law, 405
Polarizability, 60, 61, 164, 168
Polarization effect, 204
Polarized light, 46, 161, 162, 163
Polaron, 215
Polyampholytes, 91
Polycondensation, 27, 40, 63
Polyelectrolytes, 91-4, 410, 411
Polyethylene, 63, 74, 75
Polyglutamic acid, 109
Polymerases, 54, 284, 285
Polymerization, 63
Polymethacrylic acid, 91
Polypeptide chain, 27, 34, 153
  conformations of, 97-103, 240
Polypeptides, 34, 54, 95, 153, 156, 567

Polyribosome, 282
Polysaccharides, 49, 50, 133, 200
Polysome, 282, 294
Population genetics, 18
Porphyrins, 52, 53, 152, 227
Positional information, 590, 591
Potential barrier heights, 67
Potential energy, 66 *et seq.*, 74
Presumptive eye, 609
Prey-predator model, 515, 534
Prigogine-Lefever-Nicolis model, 521-3, 543
Prigogine theorem, 330, 332, 340, 502
Primary structure, 119
Primordial atmosphere, 27
Processing, 311
Procollagen, 138
Proteases, 35
Protein globule, 114-9, 210
Proteinoids, 567, 579
Protein(s), 26, 27 *et seq.*, 46, 54, 97, 133, 406
  as enzymes, 188 *et seq.*
  as polyelectrolytes, 91
  biosynthesis, 36, 43, 45, 119, 275, 280 *et seq.*
  composition of, 34, 35
  contractile, 95, 411, 416, 437, 439
  crystalline, 145 *et seq.*
  denaturation, 96, 117, 128-32, 145, 160
  fibrous (fibrillar), 100, 101, 138-41, 413
  functions of, 95, 96, 124
  globular, 91, 96, 114 *et seq.*, 125, 146
  hierarchy of structures, 120
  hypochromism in, 156
  in membranes, 348, 349, 350
  lipo-, 404
  muscle, 411-7
  nonhistone, 309, 310
  peptides, 97
  physics of, 95 *et seq.*
  primary structure of, 35, 36, 37, 119
  quaternary structure of, 115
  renaturation of, 119
  secondary structure of, 103, 114
  self-assembly of, 115, 119, 121, 122, 313
  stability of, 125 *et seq.*
  structural, 95, 96
  synthesis of, 280 *et seq.*
  three-dimensional structure of, 114, 119 *et seq.*
  tertiary structure of, 114, 119 *et seq.*
  transcription of, 280
  ultracentrifugation of, 86
Proteolytic hydrolysis, 35
Protomers, 354
Protonmotive force, 450
Proton pump, 453, 495, 498
Protoplasm movement, 433
Proximity effect, 204, 205
Puffs, 310
Purine bases, 39, 40
Pyridoxal phosphate, 171, 200, 202, 203
Pyrimidine bases, 39, 40
Pyruvate, 441, 442, 443
Pyruvic acid, 56, 441

Quality factor, 569, 571
Quantum chemistry, 59, 216
Quantum fluctuations, 489
Quantum mechanics, 11, 15
Quantum yield
  luminescence, 158
  photosynthesis, 471, 473

Rabin reaction, 223
Racemic mixture, 45 *et seq.*
Rack model, 210, 211
Radiationless transitions, 157, 159
Raman scattering, 178, 179, 180
Reaction coordinate, 189, 321
Reading frame, 277, 278
Recapitulation, 588, 609
Red blood cells (erythrocytes), 90
Red drop, 471
Redox potential, 443, 466
Redundancy, 604
Redundant information, 604 *et seq.*
Reflection coefficient, 357
Refractive index, 164
Refractory period (state), 381, 382, 385, 559-62
Regulator gene, 306
Relaxation
  chemical, 223, 224
  spin-lattice, 184
  spin-spin, 184
  time, 107, 137, 184, 335
Relaxation conformational transitions, 458
Renaturation, 119, 265
Resilin, 439
Resonance Raman scattering, 182
Respiration, 53, 56, 57, 58, 348, 469, 539
Respiratory chain, 57, 58, 440 *et seq.*
Resting membrane potential, 382, 385, 387, 395

Restriction, 286
Restriction enzymes, 286
Retina, 485, 486, 487, 488
Retinal, 52, 490-2
Retinol, 490
Reverberators, 545, 546, 560-3
Reverse transcriptase, 285, 286
Reverse transcription, 237, 280
Reversible thermodynamics, 21
Reversion, 276, 277
Revertant, 276, 277, 278
Revertase, 285, 286
Rheobase, 381
Rhodopsin, 348, 351, 352, 355, 490-7
Ribonucleic acid (RNA), 38, 42
  base pairing in, 246, 247
  double-helical portions of, 247
  intramolecular interactions in, 247
  polymerase, 284, 285, 286, 310
  ribosomal, 281
  structure of, 246, 247, 304
  synthesis of, 308, 310
  types of, 42, 43
Ribose, 38, 40, 41, 240
Ribosomal RNA, 42, 281, 290, 291
Ribosomes, 237, 280, 281, 289-94
RNA polymerase, 284, 285, 286, 310
Rods and cones, 486, 487
Rotamers, 70, 75
Rotational-isomeric theory, 71 *et seq.*
Rotational isomerism, 70 *et seq.*
Rotational strength, 164 *et seq*,
Rubber elasticity, 64, 65, 73
Rubredoxin, 152

Saddle, 510, 512
Saltatory mechanism, 386
Sarcomere, 411, 412, 413
Sarcoplasm, 418
Sarcoplasmic reticulum, 411, 418
Scatchard graph, 267
Scattering function, 90
Scattering indicatrix, 149
Schiff base, 200, 498
Schwann cells, 377, 378
Screening constant, 185
Second law of thermodynamics, 19, 20, 320
Sedimentation, 23, 85-9
  coefficient, 87
  constant, 87
  rate, 87
Selective value, of information, 570, 576
Self-assembly, 115, 119, 121, 122, 313
Self-organization, 572 *et seq.*
Semiconservative model, 243
Sensory neuron, 377
Separatrices, 510
Shannon formula, 315, 604
Shpolsky line spectra, 157
Sickle-cell anemia, 36
Silent mutation, 301
Silk fibroin, 141
Skeletal muscles, 411 *et seq.*
Sliding filament model, 416, 424
Smell receptors, 372-5
Sodium pump, 362-5, 399, 400, 401
Soliton, 423
Spacers, 311
Speciation, 614
Spectropolarimetry, 168
Sperm whale myoglobin, 36, 37, 144
Sphingomyelin, 51
Spike, 379
Spin-label techniques, 187
Spin-lattice relaxation, 184
Spin-spin interaction, 184
Spiral waves, 545, 546, 560-63
Splitting factor, 183
Squid axon, 380, 381, 382, 383, 387
Stable focus, 511, 512
Stable node, 510, 511, 512
Stable systems, 512, 514
Stacking interactions, 248, 249
Starch, 50, 57, 207, 552
Static order, 340
Stationary state, 20, 341
Stereochemical theory, 120, 121, 373
Steroid hormones, 54
Stochastic matrix, 588
Stochastic processes, 504, 532-5, 588
Stokes' law, 89
Stress-strain relation, 64
Stretched state (rack), 210, 211
Strong interactions, 58
Submitochondrial particles, 178, 445, 454, 455, 458
Substrate, 192 *et seq.*
Sucrose, 49
Sugar, 38, 240, 241
Superhelix, 309
Supermolecular structures, 96, 309, 347, 406, 535
Suppressor cells, 598
Svedberg formula, 87, 88
Svedberg unit, 87
Symport, 453
Synapse, 377, 378
Synaptic cleft, 402, 404
Synaptic transmission, 402-5
Synchrotron radiation, 150, 433
Syncytium, in muscle, 388, 394
Synergetics, 345

Taste reception, 375, 376
Tautomerism, 39
Taxons, 617
Template replication, 268
Teorell membrane oscillators, 555
Testosterone, 54, 55
Tetanus, 421
Texture, 147
Thaumatin, 375
Theory of catastrophes, 512
Theory of control, vision, 485
Theory of graphs, 223, 345
Theory of systems, 536
Thermal diffusion, 323
Thermodynamic branch, 344
Thermodynamic chain flexibility, 76
Thermodynamics, 13, 16, 19-22
Theta point, 84
Thiamine, phosphate, 52
Thylakoid disk, 347, 477, 479, 481
Thyroglobulin, 54
Thyroid gland, 54
Thyroid hormones, 54
Thyroxin, 54
Titration, 32
Tobacco mosaic virus, 115, 279
Totipotency, 609
Transamination, 200, 202
Transcription, 95, 237, 280, 308
Transcriptons, 311, 312
Transduction, 616
Transfer factors, 292
Transfer RNA (tRNA), 42, 281 *et seq.*
  aminoacyl, 292, 293
  clover-leaf model of, 287, 288, 289
  in protein biosynthesis, 292-294
  peptidyl, 292, 293
  phenylalanine, 289
  secondary structure of, 286, 287, 288
  tertiary structure of, 289
Transition dipole moment, 155, 156
Transition-state theory, 189
Translation, 95, 282 *et seq.*
Translocation, 293
Transmission coefficient, 191
Transmitter substance, 402
Transposons, 616
Transversion, 303
Tricarboxylic acid cycle, 441 *et seq.*
Tropocollagen, 138, 139
Tropomyosin, 415
Troponin, 415, 418
Trypsin, 35
Tunnel effect, 459
Tunnelling transition, 458
Turbidity coefficient, 89
Turing bifurcation, 524
Turnover number, 198
Twitch, 421

Ultracentrifugation, 85, 86
Ultrasound, 128
Uncoupler, 449
Unipotency, 609
Unit membrane model, 348, 349, 350
Universe, origin of, 16
Unstable focus, 511, 512
Unstable node, 510, 511
Unstable systems, 512

Valence-optical scheme, 179
Valinomycin, 369, 370, 371
van der Waals equation, 79, 529
van der Waals forces, 59-61, 68, 115
van der Waals radii, 101, 104, 229
van't Hoff law, 85
Variability, 12, 17, 18
Vasopressin, 54
Vavilov-Feofilov theory, 158
Vavilov law, 158
Vegetable oils, 50
Virial coefficients, 85
Viscosimetry, 85
Viscosity, 86, 91
Vision, 484-9
Visual pigments, 490
Visual purple, 490
Vital force, 13
Vitalism, 13
Vitamins, 51, 52, 490
Voigt model, 432
Voltage-clamp technique, 389
Volume effects, 83, 84, 91
Vulnerable zone, 562

Water, 29, 115, 116
  photodissociation of, 565
  structure of, 106-8
Watson-Crick model, 248
Watson-Crick pairs, 239, 243, 263, 269, 276
Weak interactions, 58-62
White box, 536
Wild type, 276, 277

X-ray diffraction, 23, 107, 125, 142 *et seq.*, 229

Zeeman splitting, 183
Zero-point energy, 60, 61
Zimm method, 90
Zone electrophoresis, 93
Zoology, 12, 17
Zwitterion, 27
Zygote, 15, 18

Printed in the Union of Soviet Socialist Republics